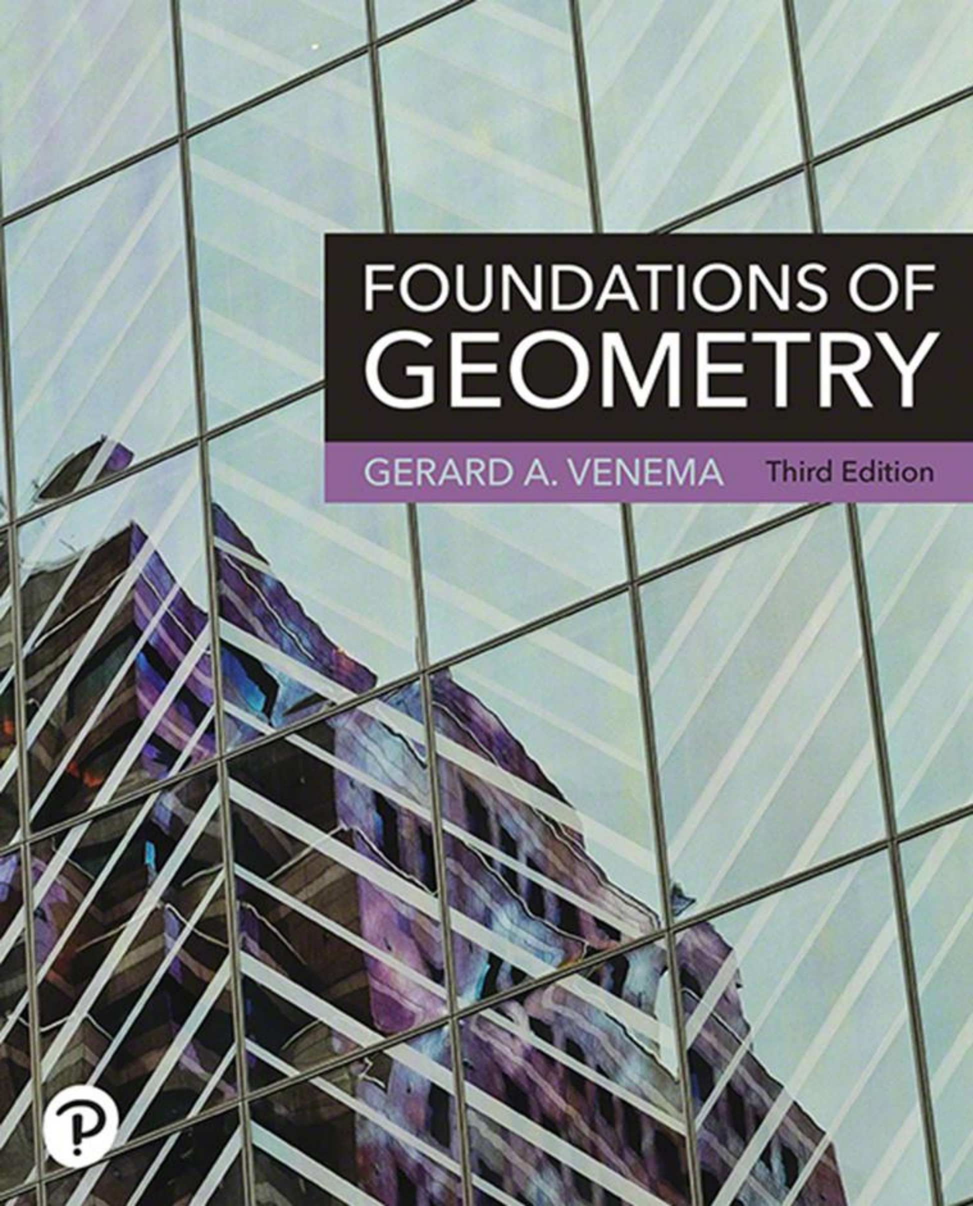
FOUNDATIONS OF GEOMETRY

GERARD A. VENEMA

Third Edition



Third Edition

Foundations of Geometry

Gerard A. Venema

Department of Mathematics and Statistics
Calvin University

Pearson

Content Management: Jon Krebs, Jeff Weidenaar
Content Production: Erin Carreiro, Ron Hampton
Product Management: Chelsea Kharakozova
Product Marketing: Alicia Wilson
Rights and Permissions: Tanvi Bhatia
Composition: Paul C. Anagnostopoulos, Windfall Software
Cover Image: Wilson Tan/EyeEm/Getty Images

Please contact https://support.pearson.com/getsupport/s/ with any queries on this content.

Credits:
P. 384, CRICKET MEDIA INCORPORATED: Reprinted from David Hilbert, *Foundations of Geometry,* Open Court Publishing Company, La Salle, Illinois, 1971 by permission of the Open Court Publishing Company.
P. 386, PRINCETON UNIVERSITY: Birkhoff, George D., "A set of postulates for plane geometry," *Annals of Mathematics* 33: pp. 329–345.
P. 389, School Mathematics Study Group: School Mathematics Study Group.
P. 391, University of Chicago School Mathematics Project: From UCSMP Geometry © 2016, The University of Chicago, pp. S1–S3.

Library of Congress Cataloging-in-Publication Data
is available on file at the Library of Congress.

ScoutAutomatedPrintCode

eText Access Card:
ISBN-10: 0136845126
ISBN-13: 9780136845126

Print Rental:
ISBN-10: 0136845266
ISBN-13: 9780136845263

Pearson

To Patricia

This page intentionally left blank

Contents

This page intentionally left blank

Preface

This is a textbook for an undergraduate course in geometry. The text is targeted at mathematics students who have completed the calculus sequence, and perhaps a first course in linear algebra, but who have not necessarily encountered such upper-level mathematics courses as real analysis or abstract algebra. A course based on this book will enrich the education of all mathematics majors and will ease their transition into more advanced mathematics courses. The book includes emphases that make it especially appropriate as the textbook for a geometry course taken by future high school mathematics teachers.

What is distinctive about this book

What distinguishes this textbook from other undergraduate geometry textbooks is the foundational approach. The book attempts to present a traditional axiomatic treatment of geometry while at the same time communicating the intuitive beauty and fascination of the subject. The book gives a complete, rigorous development of plane geometry, and does so in a way that leads students to appreciate the power of proofs and the feeling of satisfaction that comes with the ability to write good proofs. Because of the emphasis on proof-writing, the course can serve as a bridge course in the mathematics major and can be the course in which students learn to write proofs. Another distinguishing characteristic of the book is the fact that the needs of future high school teachers are always considered. Even though attention is paid to the requirements of future teachers, a geometry course based on this book can be an important component of the education of any mathematics student because it is one of the few places in which a student will see a complete, axiomatic development of a branch of mathematics.

What is new in the third edition

- There is now an eText edition of the book. The eText includes interactive figures that allow users to explore various constructions and to make discoveries for themselves.
- Some material was rewritten to bring it more in line with current standards.
- All typographical errors that have been brought to the author's attention were corrected.
- The presentation of the material in several sections was rethought in light of comments from users of the second edition.
- More exercises have been added throughout.

The various themes and emphases of the book as well as the supplementary material included in the eText are described in more detail in the following paragraphs.

The foundations of geometry

A principal goal of the text is to study the *foundations* of geometry. That means returning to the beginnings of geometry, exposing exactly what is assumed there, and building the entire subject on those foundations. Such careful attention to the foundations has a long tradition in geometry, going back more than two thousand years to Euclid and the ancient Greeks. Over the years since Euclid wrote his famous *Elements*, there have been profound changes in the way in which the foundations have been understood. Most of those changes have been by-products of efforts to understand the true place of Euclid's parallel postulate in the foundations, so the parallel postulate is one of the primary emphases of this book.

Proofs

A secondary goal of the text is to teach the art of writing proofs. There is general recognition of the need for a course in which mathematics students learn how to write good proofs. Such a course should serve as a bridge between the lower-level mathematics courses, which are largely technique oriented, and the upper-level courses, which tend to be much more conceptual. This book uses geometry as the vehicle for helping students to write and appreciate proofs. The ability to write proofs is a skill that can be acquired only by actually practicing it, so most of the material on writing proofs is integrated into the text and the attention to proof permeates the entire text. This means that the book can also be used in classes where the students already have experience writing proofs; despite the emphasis on writing proofs, the book is still primarily a geometry text.

Having the geometry course serve as the introduction to proof represents a return to tradition in that the course in Euclidean geometry has for thousands of years been seen as the standard introduction to logic, rigor, and proof in mathematics. Using the geometry course this way makes historical sense because the axiomatic method was first introduced in geometry and geometry remains the branch of mathematics in which that method has had its greatest success. While proof and logical deduction are still emphasized in the standards for high school mathematics, most high school students no longer take a full-year course devoted exclusively to geometry with a sustained emphasis on proof. This makes it more important than ever that we teach a good college-level geometry course to all mathematics students. By doing so we can return geometry to its place as the subject in which students first learn to appreciate the importance of clearly spelling out assumptions and deducing results from those assumptions via careful logical reasoning.

The emphasis on proof makes this text a do-it-yourself course in that the reader will be asked to supply proofs for many of the key theorems. Students who diligently work the exercises come away from a course based on this book with a sense that they have an unusually deep understanding of the material. In this way the student will not only learn the mechanics of good proof writing style but should also come to more fully appreciate the important role proof plays in an understanding of mathematics.

Historical and philosophical perspective

A final goal of the text is to present a historical perspective on geometry. Geometry is a dynamic subject that has changed over time. It is a part of human culture that was created and developed by people who were very much products of their time and place. The foundations of geometry have been challenged and reformulated over the years, and beliefs about the relationship between geometry and the real world have been challenged as well.

The material in the book is presented in a way that is sensitive to such historical and philosophical issues. This does not mean that the material is presented in a strictly historical order or that there are lengthy historical discussions but rather that geometry is presented in such a way that the reader can understand and appreciate the historical development of the subject and so that it would be natural to investigate the history of the subject while learning it. Many chapters include suggested readings on the history of geometry that can be used to enrich the text.

Throughout the book there are references to philosophical issues that arise in geometry. For example, one question that naturally occurs to anyone studying non-Euclidean geometry is this: What is the connection between the abstract entities that are studied in a course on the foundations of geometry and properties of physical space? The book does not present dogmatic answers to such questions, but instead simply raises them in an effort to promote student thinking. The hope is that this will serve to counter the common perception that mathematics is a subject in which every question has a single correct answer and in which there is no room for creative ideas or opinions.

Technology

In recent years powerful computer software has been developed that can be used to explore geometry. The study of geometry from this book can be greatly enhanced by such dynamic software and the reader is encouraged to find appropriate ways in which to incorporate this technology into the geometry course. While software can enrich the experience of learning geometry from this book, its use is not required. The book can be read and studied quite profitably without it.

The author recommends the use of the dynamic mathematics software program GeoGebra. GeoGebra is free software that is intended to be used for teaching and learning mathematics. It may be downloaded from the website www.geogebra.org. The software has many great features that make it ideal for use in the geometry classroom, but the main advantage it has over commercial geometry software is the fact that it is free and runs under any of the standard computer operating systems. This means that students can load the program on their own laptops and will always have access to it.

In the first part of the text (Chapters 2 through 4), the objective is to carefully expose all the assumptions that form the foundations of geometry and to understand for ourselves how the basic results of geometry are built on those foundations. For most users, the software is a black box in the sense that we either don't know what assumptions are built into it or we have only the authors' description of what went into the software. As a result, software is of limited use in that part of the course and it will not be mentioned explicitly in the first four chapters of the book. But ideally readers should be using it to draw diagrams and to experiment with what happens when they vary the data in the theorems. During that phase of the course the main function of the software is to illustrate one possible interpretation of the relationships being studied.

It is in the second half of the course that the software comes into its own. Computer software is ideal for experimenting, exploring, and discovering new relationships. In order to illustrate that, several of the later chapters include sections in which the software is used to explore ideas that go beyond those that are presented in detail and to discover new relationships. In particular, there are such exploratory sections in the chapters on Euclidean geometry and circles. The entire chapter on constructions is written as an exploration with only a limited number of proofs or hints provided in the text. The exploratory sections of the text have been expanded into a short book entitled *Exploring Advanced Euclidean Geometry with GeoGebra* [Ven13].

The eText

While the printed text assumes readers are working directly in GeoGebra to construct their own sketches, the eText provides a number of GeoGebra-based interactive figures that are ready to be used for exploration. In each interactive figure, users are instructed to click on boxes that reveal in stages how the figure is constructed. Users are then guided through a dynamic exploration of the geometric object under study. In most cases users are asked questions that are intended to prompt exploration of various aspects of the figure. After users have had a chance to explore the answer to a question for themselves, they can click on a box to reveal the textbook answer.

Many of the interactive figures illustrate results of more advanced Euclidean geometry that are both surprising and beautiful. The interactive figures facilitate an understanding of what the theorems say and allow users to see some of the amazing relationships for themselves. The author hopes this will help users to appreciate the elegance and beauty of Euclidean geometry and to better understand why the subject has captivated the interest of so many people over the past two thousand years. Other interactive figures allow the user to explore hyperbolic geometry by working with various objects in the Poincaré disk model.

National standards

A geometry course based on this textbook will be consistent with the most recent recommendations from relevant professional associations. In particular, the text follows the recommendations in the *2015 Curriculum Guide of the Committee on the Undergraduate Program in Mathematics* [CUP15]. In addition, since a significant portion of the audience for an undergraduate geometry course consists of future high school geometry teachers, the book implements current national standards regarding the mathematical education of teachers. Those standards are contained in the updated report *The Mathematical Education of Teachers II* [CBM12] and the *Common Core State Standards for Mathematics* [CCS10].

The Mathematical Association of America maintains a standing Committee on the Undergraduate Program in Mathematics (CUPM). Roughly once each decade the committee publishes a comprehensive curriculum guide that sets standards for the undergraduate mathematics major. The most recent guide was published in 2015. The abridged printed version [CUP15] contains guidelines for the mathematics major and certain key courses. The expanded online version, which includes reports on individual courses, is found at https://www.maa.org/node/790342. A subcommittee of the CUPM, chaired by the author of this book, wrote the report on the undergraduate geometry course. The committee described several possible syllabi for such a course and this textbook follows the recommended syllabus for a course in axiomatic geometry. While the geometry subcommittee was open to a number of possible syllabi and emphases in the geometry course, it recommended that every undergraduate geometry course include a study of transformations. For that reason the chapter on transformations in this book is written in a way that allows any course taught from the book to include a study of transformations.

The report on *The Mathematical Education of Teachers* (MET) was originally written in 2001 and was updated in 2012. The principal recommendation of MET is that "Prospective teachers need mathematics courses that develop a deep understanding of the mathematics they will teach" [CBM12, page 17]. This text is designed to do precisely that in the area of geometry. The book makes a conscious effort to ensure that there are clear connections between the geometry in the course and the geometry that future high school instructors will teach. An example of the way in which connections with high school geometry have influenced the design of the text is the choice of the axioms that are used as the starting point. One of the main goals of the text is to help preservice teachers understand the logical foundations of the geometry course they will teach, and that goal can best be accomplished in the context of axioms that are like the ones they will encounter later in the classroom. Thus the axioms on which the development of the geometry in the text is based are as close as possible to those that are used in contemporary high school textbooks. In addition, as there is no standard set of axioms that is common to all such high school geometry courses, various axiom systems are considered in an appendix and the merits and advantages of each are discussed. This close attention to the statements of the axioms is just one example of the many connections with high school geometry that are brought in as the course progresses.

One of the recurring themes in MET is the recommendation that prospective teachers must acquire an understanding of high school mathematics that goes well beyond that of a typical high school graduate. One way in which such understanding of geometry is often measured is in terms of the van Hiele model of geometric thought. This model is described in Appendix D. The goal of most high school courses is to develop student thinking to Level 3. A goal of this text is to bring students to Level 4 (or to Level 5, depending on whether the first level is numbered 0 or 1). It is recognized, however, that not all students entering the course are already at Level 3 and so the early part of the text is designed to ensure that students are brought to that level first.

The *Common Core State Standards for Mathematics* (CCSSM) specify what should be included in the high school geometry curriculum. This book attempts to give future teachers a grounding in the themes and perspectives described there. In particular, there is an emphasis on Euclidean geometry and the parallel postulate. The transformational approach to congruence and similarity, the approach that is promoted by CCSSM, is studied in Chapter 10 and is related there to other, more traditional, ways of interpreting congruence and similarity. In fact, all of the specific topics listed in CCSSM are covered in the text. Finally, CCSSM mentions that ". . . in college some students will develop Euclidean and other geometries carefully from a small set of axioms" [CCS10, page 74]. As detailed earlier, a course based on this textbook is exactly the kind of course envisioned in that remark.

Organization of the book

The book begins with a brief look at Euclid's *Elements*, and Euclid's method of organization is used as motivation for the concept of an axiomatic system. A system of axioms for geometry is then carefully laid out. The axioms used here are based on the real numbers, in the spirit of Birkhoff, and their statements have been kept as close to those in contemporary high school textbooks as is possible.

After the axioms have been stated and certain foundational issues faced, neutral geometry, in which no parallel postulate is assumed, is extensively explored. Next both Euclidean and hyperbolic geometries are investigated from an axiomatic point of view. In order to get as quickly as possible to some of the interesting results of non-Euclidean geometry, the first part of the book focuses exclusively on results regarding lines, parallelism, and triangles. Only after those topics have been treated separately in neutral, Euclidean, and hyperbolic geometries are results on area, circles, and construction introduced. While the treatment of these subjects does not exactly follow Euclid, it roughly parallels Euclid in the sense that Euclid collected most of his propositions about area in Book II and most of his propositions about circles in Books III and IV. The three chapters covering area, circles, and construction complete the coverage of the major theorems of Books I through VI of the *Elements*.

The more modern notion of a transformation is introduced next and some of the standard results regarding transformations of the plane are explored. A complete proof of the classification of the rigid motions of both the Euclidean and hyperbolic planes is included. There is a discussion of how the foundations of geometry can be reorganized to reflect the transformational point of view (as is common practice in contemporary high school geometry textbooks). Specifically, it is possible to replace the Side-Angle-Side Postulate with a postulate that asserts the existence of certain reflections.

The standard models for hyperbolic geometry are carefully constructed and the results of the chapter on transformations are used to verify their properties. The chapter on models can be relatively short because all the hard technical work involved in the constructions is done in the preceding chapter. The final chapter includes a study of some of the polygonal models that have recently been developed to help students understand what it means to say that hyperbolic space is negatively curved. The book ends with a discussion of the practical significance of non-Euclidean geometry and a brief look at the geometry of the real world.

Designing a course

A full-year course should cover essentially all the material in the text. There can be some variation based on instructor and student interest, but most or all of every chapter should be included.

An instructor teaching a one-semester or one-quarter course will be forced to pick and choose. It is important that this be done carefully so that the course reaches some of the interesting and useful material that is to be found in the second half of the book.

- Chapter 1 sets the stage for what is to come, so it should be covered in some way. But it can be discussed briefly in class and then assigned as reading.
- Chapter 2 should definitely be covered because it establishes the basic framework for the treatment of geometry that follows.
- The basic coverage of geometry begins with Chapter 3. Chapters 3 and 4 form the heart of a one-semester course. Those chapters should be included in any course taught from the book.
- At least some of Chapters 5 and 10 should also be included in any course.
- Starting with Chapter 7, the chapters are largely independent of each other and an instructor can select material from them based on the interests and needs of the class.

Several sample course outlines are included below. Many other variations are possible. It should be noted that the suggested outlines are ambitious and many instructors will choose to cover less.

A course emphasizing Euclidean geometry.

Chapter	Topic	Number of weeks
1 & 2	Preliminaries	≤ 2
3	Axioms	2
4	Neutral geometry	3
5	Euclidean geometry	2
7	Area	1–2
8	Circles	1–2
10	Transformations	2

A course emphasizing non-Euclidean geometry.

Chapter	Topic	Number of weeks
1 & 2	Preliminaries	≤ 2
3	Axioms	2
4	Neutral geometry	3
5	Euclidean geometry	1
6	Hyperbolic geometry	2
7	Area	1–2
10	Transformations	1
11	Models	1–2
12	Geometry of space	1

A course for future high school teachers.

Chapter	Topic	Number of weeks
1 & 2	Preliminaries	≤ 2
3	Axioms	2
4	Neutral geometry	3
5	Euclidean geometry	1
6	Hyperbolic geometry	1
7	Area	1
8	Circles	1
10	Transformations	1
11	Models	1
12	Geometry of space	1

The suggested course for future high school teachers includes just a brief introduction to each of the topics in later chapters. The idea is that the course should provide enough background so that students can study those topics in more depth later if they need to. It is hoped that this book can serve as a valuable reference for those who go on to teach geometry courses. The book could be a resource that provides information about rigorous treatments of such topics as parallel lines, area, circles, constructions, transformations, and so on, that are part of the high school curriculum.

Instructor's Solutions Manual

There is an *Instructor's Solutions Manual* that contains solutions to all the exercises as well as additional information on teaching from the book. Instructors can obtain the manual from the Pearson Higher Education website.

Acknowledgments

I once again thank all those who were acknowledged in the first two editions of the book.

Numerous users of the second edition contacted me to share observations, questions, and suggestions. I received so many contributions that it is not possible to acknowledge all of them individually, but every one of the comments is appreciated and all were carefully considered when the third edition was being prepared.

Even though I cannot list all those who contributed to this edition, three individuals who made substantial contributions to the content should be mentioned. Dan Velleman, Amherst College, provided extensive feedback that influenced many of the revisions in this edition, especially the treatment of the axioms in Chapter 3. Harald Hanche-Olsen, Norwegian University of Science and Technology, suggested a different approach to transformations that greatly influenced the new treatment of rotations in Section 10.2. Jerry Grossman, Oakland University, read the text carefully and brought a large number of errors to my attention.

I also wish to thank all those who have assisted with the production of the third edition. In particular, Paul Anagnostopoulos, Windfall Software, did excellent work on the typesetting. MaryEllen Oliver did an unusually careful and thorough job of proofreading. Ron Weickart, Network Graphics, produced the figures. Alicia Hofstätter, GeoGebra, provided valuable advice regarding the design of the interactive figures. John Samons, Florida State University at Jacksonville, did the accuracy checking. Ron Hampton and Jonathan Krebs at Pearson Education assisted with the content development. Jeff Weidenaar, Content Manager at Pearson, approved the project and oversaw its progress.

Finally, I want to thank my wife Patricia and my entire extended family for their loving support throughout the entire process of writing and revising this book.

Gerard A. Venema
Calvin University
September 2020

CHAPTER 1

Prologue: Euclid's *Elements*

Our study of geometry begins with an examination of the historical origins of the axiomatic method in geometry. While the material in this chapter is not a mathematical prerequisite for what comes later, an appreciation of the historical roots of axiomatic thinking is essential to an understanding of why the foundations of geometry are systematized as they are.

1.1 GEOMETRY BEFORE EUCLID

Geometry is an ancient subject. Its roots go back thousands of years and geometric ideas of one kind or another are found in nearly every human culture. The beauty of geometric patterns is universally appreciated and often investigated in an informal way. The systematic study of geometry as we know it emerged more than 4000 years ago in Mesopotamia, Egypt, India, and China.

Because the Nile River annually flooded vast areas of land and obliterated property lines, surveying and measuring were important to the ancient Egyptians. This practical interest may have motivated their study of geometry. Egyptian geometry was mostly an empirical science, consisting of many rule-of-thumb procedures that were arrived at through experimentation, observation, and trial and error. Formulas were approximate ones that appeared to work, or at least gave answers that were close enough for practical purposes. But the ancient Egyptians were also aware of more general principles, such as special cases of the Pythagorean Theorem and formulas for volumes.

The ancient Mesopotamians, or Babylonians, had an even more advanced understanding of geometry. They knew the Pythagorean Theorem long before Pythagoras. They discovered some of the area-based proofs of the theorem that will be discussed in Chapter 7, and knew a general method that generates all triples of integers that are lengths of sides of right triangles. In India, ancient texts apply the Pythagorean Theorem to geometric problems associated with the design of structures. The Pythagorean Theorem was also discovered in China at roughly the same time.

About 2500 years ago there was a profound change in the way geometry was practiced: Greek mathematicians introduced abstraction, logical deduction, and proof into geometry. They insisted that geometric results be based on logical reasoning from first principles. In theory this made the results of geometry exact, certain, and undeniable, rather than just likely or approximate. It also took geometry out of the realm of everyday experience and made it a subject that studies abstract entities. Since the purpose of this course is to study the logical foundations of geometry, it is natural that we should start with the geometry of the ancient Greeks.

The process of introducing logic into geometry apparently began with Thales of Miletus around 600 BC and culminated in the work of Euclid of Alexandria in approximately 300 BC. Euclid is the most famous of the Greek geometers and his name is still universally associated with the geometry that is studied in schools today. Most of the ideas that are included in what we call "Euclidean Geometry" probably did not originate with Euclid himself; rather, Euclid's contribution was to organize and present the results of Greek geometry in a logical and coherent way. He published his results in a series of thirteen books known as his *Elements*. We begin our study of geometry by examining those *Elements* because they set the agenda for geometry for the next two millennia and more.

1.2 THE LOGICAL STRUCTURE OF EUCLID'S *ELEMENTS*

Euclid's *Elements* are organized according to strict logical rules. Euclid begins each book with a list of definitions of the technical terms he will use in that book. In Book I he next states five "postulates" and five "common notions." These are assumptions that are meant to be accepted without proof. Both the postulates and common notions are basic statements whose truth should be evident to any reasonable person. They are the starting point for what follows. Euclid recognized that it is not possible to prove everything, that he had to start somewhere, and he attempted to be clear about exactly what his assumptions were.

Most of Euclid's postulates are simple statements of intuitively obvious and undeniable facts about space. For example, Postulate I asserts that it is possible to draw a straight line through any two given points. Postulate II says that a straight line segment can be extended to a longer segment. Postulate III states that it is possible to construct a circle with any given center and radius. Traditionally these first three postulates have been associated with the tools that are used to implement them on a piece of paper. The first two postulates allow two different uses of a straightedge: A straightedge can be used to draw a line segment connecting any two points or to extend a given line segment to a longer one. The third postulate affirms that a compass can be used to construct a circle with a given center and radius. Thus the first three postulates simply permit the familiar straightedge and compass constructions of high school geometry.

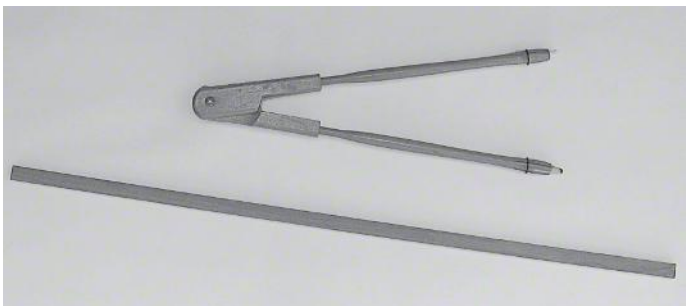


FIGURE 1.1 Euclid's tools: a compass and a straightedge

The fourth postulate asserts that all right angles are congruent ("equal" in Euclid's terminology). The fifth postulate makes a more subtle and complicated assertion about two lines that are cut by a transversal. These last two postulates are the two technical facts about geometry that Euclid needs in his proofs.

The common notions are also intuitively obvious facts that Euclid plans to use in his development of geometry. The difference between the common notions and the postulates is that the common notions are not peculiar to geometry but are common to all branches of mathematics. They are everyday, common-sense assumptions. Most spell out properties of equality, at least as Euclid used the term *equal*.

The largest part of each book of the *Elements* consists of propositions and proofs. These too are organized in a strict, logical progression. The first proposition is proved using only the postulates, Proposition 2 is proved using only the postulates and Proposition 1, and so on. Thus the entire edifice is built on just the postulates and common notions; once these are granted, everything else follows logically and inevitably from them. What is astonishing is the number and variety of propositions that can be deduced from so few assumptions.

1.3 THE HISTORICAL SIGNIFICANCE OF EUCLID'S *ELEMENTS*

It is nearly impossible to overstate the importance of Euclid's *Elements* in the development of mathematics and human culture generally. From the time they were written, the *Elements* have been held up as the standard for the way in which careful thought ought to be organized. They became the model for the development of all scientific and philosophical theories. What was especially admired about Euclid's work was the way in which he clearly laid out his assumptions and then used pure logic to deduce an incredibly varied and extensive set of conclusions from them.

Up until the twentieth century, Euclid's *Elements* were the textbook from which all students learned both geometry and logic. Even today the geometry in school textbooks is presented in a way that is remarkably close to that of Euclid. Furthermore, much of what mathematicians did during the next two thousand years centered around tying up loose ends left by Euclid. Countless mathematicians spent their careers trying to solve problems that were raised by Greek geometers of antiquity and trying to improve on Euclid's treatment of the foundations.

Most of the efforts at improvement focused on Euclid's Fifth Postulate. Even though the statement does not explicitly mention parallel lines, this postulate is usually referred to as "Euclid's Parallel Postulate." It asserts that two lines that are cut by a transversal must intersect on one side of the transversal if the interior angles on that side of the transversal sum to less than two right angles. In particular, it proclaims that the condition on the angles formed by a transversal implies that the two given lines are not parallel. Thus it is really a statement about nonparallel lines. As we shall see later, the postulate can be reformulated in ways that make it more obviously and directly a statement about parallel lines.

A quick reading of the postulates (see the next section) reveals that Postulate V is noticeably different from the others. For one thing, its statement is much longer than those of the other postulates. A more significant difference is the fact that it involves a fairly complicated arrangement of lines and also a certain amount of ambiguity in that the lines must be "produced indefinitely." It is not as intuitively obvious or self-evident as the other postulates; it has the look and feel of a proposition rather than a

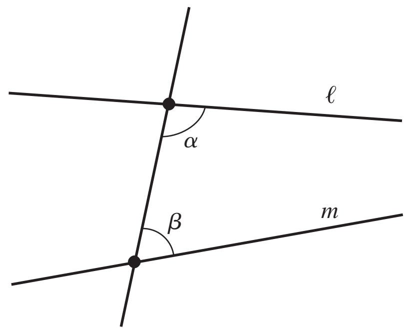


FIGURE 1.2 Euclid's Postulate V: If the sum of α and β is less than two right angles, then ℓ and m must eventually intersect

postulate. For these reasons generations of mathematicians tried to improve on Euclid by attempting to prove that Postulate V is a logical consequence of the other postulates or, failing that, they tried at least to replace Postulate V with a simpler, more intuitively obvious postulate from which Postulate V could then be deduced as a consequence. No one ever succeeded in proving the fifth postulate using just the first four postulates, but it was not until the nineteenth century that mathematicians fully understood why that was the case.

It should be recognized that these efforts at improvement were not motivated by a perception that there was anything wrong with Euclid's work. Quite the opposite: Thousands of mathematicians spent enormous amounts of time trying to improve on Euclid precisely because they thought so highly of Euclid's accomplishments. They wanted to take what was universally regarded as the crown of theoretical thought and make it even more wonderful than it already was!

Another important point is that efforts to rework Euclid's treatment of geometry led indirectly to progress in mathematics that went far beyond mere improvements in the *Elements* themselves. Attempts to prove Euclid's Fifth Postulate eventually resulted in the realization that, in some kind of stroke of genius, Euclid somehow had the great insight to pinpoint one of the deepest properties that a geometry may have. Not only that, but it was discovered that there are alternative geometries in which Euclid's Fifth Postulate fails to hold. These discoveries were made in the early nineteenth century and had far-reaching implications for all of mathematics. They opened up whole new fields of mathematical study; they also produced a revolution in the conventional view of how mathematics relates to the real world and forced a new understanding of the nature of mathematical truth.

The story of how Euclid's Parallel Postulate inspired all these developments is one of the most interesting in the history of mathematics. That story will unfold in the course of our study of geometry in this book. There is no other branch of mathematics or science that depends so profoundly and directly on one seminal text and it is only in the light of that story that the current organization of the foundations of geometry can be properly understood.

1.4 A LOOK AT BOOK I OF THE *ELEMENTS*

In order to give more substance to our discussion, we now take a direct look at parts of Book I of the *Elements*. All Euclid's postulates are stated below as well as selected definitions and propositions. The excerpts included here are chosen to illustrate the points that will be made in the following section. The translation into English is by Sir Thomas Little Heath (1861–1940). Heath used square brackets to set off anything that he thought had probably been added later and was not part of Euclid's original. A complete list of the definitions and propositions from Book I may be found in Appendix A.

Some of Euclid's definitions

Definition 1. A *point* is that which has no part.

Definition 2. A *line* is breadthless length.

Definition 4. A *straight line* is a line which lies evenly with the points on itself.

Definition 10. When a straight line set up on a straight line makes the adjacent angles equal to one another, each of the equal angles is *right*, and the straight line standing on the other is called a *perpendicular* to that on which it stands.

Definition 11. An *obtuse* angle is an angle greater than a right angle.

Definition 12. An *acute* angle is an angle less than a right angle.

Euclid's Postulates

Postulate I. To draw a straight line from any point to any point.

Postulate II. To produce a finite straight line continuously in a straight line.

Postulate III. To describe a circle with any center and distance.

Postulate IV. That all right angles are equal to one another.

Postulate V. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than the two right angles.

Euclid's Common Notions

Common Notion I. Things which equal the same thing are also equal to one another.

Common Notion II. If equals be added to equals, the wholes are equal.

Common Notion III. If equals be subtracted from equals, the remainders are equal.

Common Notion IV. Things which coincide with one another are equal to one another.

Common Notion V. The whole is greater than the part.

Three of Euclid's Propositions and their proofs

Proposition 1. On a given finite straight line to construct an equilateral triangle.

Let AB be the given finite straight line. Thus it is required to construct an equilateral triangle on the straight line AB. With center A and distance AB let the circle BCD be described [Post. III]; again, with center B and distance BA let the circle ACE be described

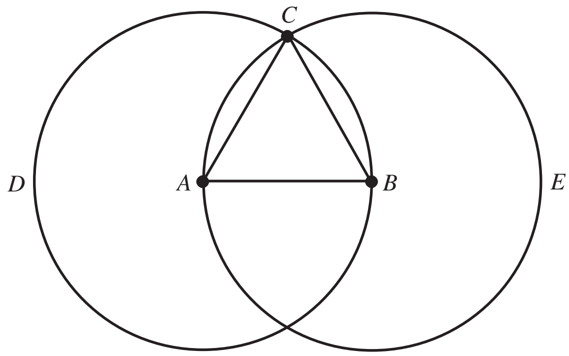


FIGURE 1.3 Euclid's diagram for Proposition 1

[Post. III]; and from the point C, in which the circles cut one another, to the points A, B let the straight lines CA, CB be joined [Post. I].

Now, since the point A is the center of the circle CDB, AC is equal to AB [Def. 15]. Again, since the point B is the center of the circle CAE, BC is equal to BA [Def. 15]. But CA was also proved equal to AB; therefore each of the straight lines CA, CB is equal to AB. And things which are equal to the same thing also equal one another [C.N. I]; therefore CA is also equal to CB. Therefore the three straight lines CA, AB, BC are equal to one another. Therefore the triangle ABC is equilateral; and it has been constructed on the given finite straight line AB.

Being what it was required to do.

Proposition 4. If two triangles have the two sides equal to two sides respectively, and have the angles contained by the equal straight lines equal, they will also have the base equal to the base, the triangle will be equal to the triangle, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend.

Let ABC, DEF be two triangles having the two sides AB, AC equal to the two sides DE, DF respectively, namely AB to DE and AC to DF, and the angle BAC equal to the angle EDF. I say that the base BC is also equal to the base EF, the triangle ABC will be equal to the triangle DEF, and the remaining angles will be equal to the remaining angles respectively, namely those which the equal sides subtend, that is, the angle ABC to the angle DEF, and the angle ACB to the angle DFE.

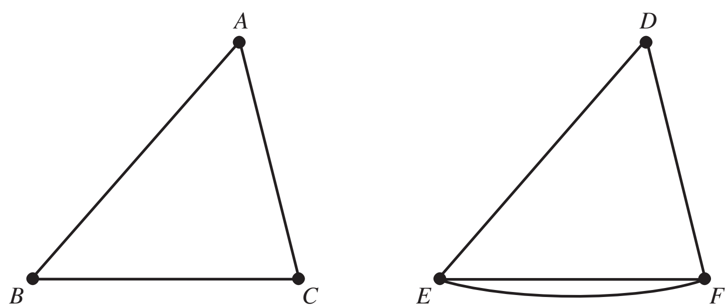


FIGURE 1.4 Euclid's diagram for Proposition 4

For, if the triangle ABC be applied to the triangle DEF, and if the point A be placed on the point D and the straight line AB on DE, then the point B will also coincide with E, because AB is equal to DE. Again, AB coinciding with DE, the straight line AC will also coincide with DF, because the angle BAC is equal to the angle EDF; hence the point C will also coincide with the point F, because AC is again equal to DF.

But B also coincided with E; hence the base BC will coincide with the base EF. [For if, when B coincides with E and C with F, the base BC does not coincide with the base EF, two straight lines will enclose a space: which is impossible. Therefore the base BC will coincide with EF] and will be equal to it [C.N. IV]. Thus the whole triangle ABC will coincide with the whole triangle DEF, and will be equal to it. And the remaining angles also coincide with the remaining angles and will be equal to them, the angle ABC to the angle DEF, and the angle ACB to the angle DFE.

Therefore etc. Being what it was required to prove.

Proposition 16. In any triangle, if one of the sides be produced, the exterior angle is greater than either of the interior and opposite angles.

Let ABC be a triangle, and let one side of it BC be produced to D; I say that the exterior angle ACD is greater than either of the interior and opposite angles CBA, BAC.

Let AC be bisected at E [Prop. 10] and let BE be joined and produced in a straight line to F; let EF be made equal to BE [Prop. 3], let FC be joined [Post. I], and let AC be drawn through to G [Post. II]. Then since AE is equal to EC, and BE to EF, the two sides AE, EB are equal to the two sides CE, EF respectively; and the angle AEB is equal to the angle FEC, for they are vertical angles [Prop. 15]. Therefore the base AB is equal to the base FC, and the triangle ABE is equal to the triangle CFE, and the remaining angles are equal to the remaining angles respectively, namely those which the equal sides subtend [Prop. 4]; therefore the angle BAE is equal to the angle ECF. But the angle ECD is greater than the angle ECF [C.N. V]; therefore the angle ACD is greater than the angle BAE.

Similarly, also if BC is bisected, the angle BCG, that is, the angle ACD [Prop. 15], can be proved greater than the angle ABC as well.

Therefore etc. Q.E.D.

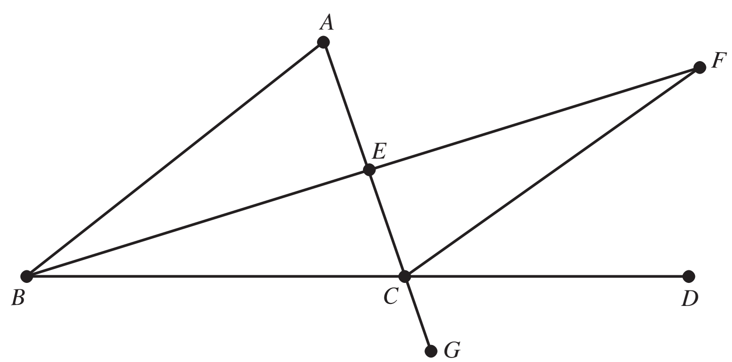


FIGURE 1.5 Euclid's diagram for Proposition 16

1.5 A CRITIQUE OF EUCLID'S *ELEMENTS*

As indicated earlier, the *Elements* have been the subject of a great deal of interest over the thousands of years since Euclid wrote them and the study of Euclid's method of organizing his material has inspired mathematicians to even greater levels of logical rigor. Originally attention was focused on Euclid's postulates, especially his fifth postulate, but efforts to clarify the role of the postulates eventually led to the realization that there are difficulties with other parts of Euclid's *Elements* as well. When the definitions and propositions are examined in the light of modern standards of rigor it becomes apparent that Euclid did not achieve all the goals he set for himself—or at least that he did not accomplish everything he was traditionally credited with having done.

Euclid purports to define all the technical terms he will use.[1] However, an examination of his definitions shows that he did not really accomplish this. The first few definitions are somewhat vague, but suggestive of intuitive concepts. An example is the very first definition, in which *point* is defined as "that which has no part." This does indeed suggest something to most people, but it is not really a rigorous definition in that it does not stipulate what sorts of objects are being considered. It is somehow understood from the context that it is only geometric objects which cannot be subdivided that are to be called points. Even then it is not completely clear what a point is: Apparently a point is pure location and has no size whatsoever. But there is nothing in the physical world of our experience that has those properties exactly. Thus we must take *point* to be some kind of idealized abstract entity and admit that its exact nature is not adequately explained by the definition. Similar comments could be made about Euclid's definitions of *line* and *straight line*.

By contrast, later definitions are more complete in that they define one technical word in terms of others that have been defined previously. Examples are Definitions 11 and 12 in which *obtuse angle* and *acute angle* are defined in terms of the previously defined *right angle*. From the point of view of modern rigor there is still a gap in these definitions because Euclid does not specify what it means for one angle to be greater than another. The difference is that these definitions are complete in themselves and would be rigorous and usable if Euclid were to first spell out what it means for one angle to be greater than another and also define what a right angle is.

Such observations have led to the realization that there are actually two kinds of technical terms. It is not really possible to define all terms; just as some statements must be accepted without proof and the other propositions proved as consequences, so some terms must be left undefined. Other technical terms can then be defined using the undefined terms and previously defined terms. This distinction will be made precise in the next chapter.

A careful reading of Euclid's proofs reveals some gaps there as well. The proof of Proposition 1 is a good example. In one sense, the proposition and its proof are simple and easy to understand. In modern terminology the proposition asserts the following: *Given two points A and B, it is possible to construct a third point C such that* $\triangle ABC$ *is an equilateral triangle.* Euclid begins with the segment from A to B. He then uses Postulate III to draw two circles of radius AB, one centered at A and the other centered at B. He takes C to be one of the two points at which the circles intersect and uses Postulate I to fill in the sides of a triangle. Euclid completes the proof by using the common notions to

1. Some scholars suggest that the definitions included in the versions of the *Elements* that have come down to us were not in Euclid's original writings, but were added later. Even if that is the case, the observations made here about the definitions are still valid.

explain why the triangle he has constructed must be equilateral. The written proof is supplemented by a diagram that makes the construction clear and convincing.

Closer examination shows, however, that Euclid assumed more than just what he stated in the postulates. In particular there is nothing explicitly stated in the postulates that would guarantee the existence of a point C at which the two circles intersect. The existence of C is taken for granted because the diagram clearly shows the circles intersecting in two points. Euclid is assuming that his diagrams accurately portray geometric relationships. That is a reasonable assumption for him to make in this context, but it is one that has not been stated in the postulates. We will see in Chapter 3 that there are unusual situations in which no point of intersection exists. In summary, Euclid is using "facts" about his points and lines that are undoubted and intuitively obvious to most readers, but which have not been explicitly stated in the postulates.

Euclid's Proposition 4 is the familiar Side-Angle-Side Congruence Condition from high school geometry. This proposition is not just a construction like Proposition 1 but asserts a logical implication: If two sides and the included angle of one triangle are congruent to the corresponding parts of a second triangle, then the remaining parts of the two triangles must also be congruent. Euclid's method of proof is interesting. He takes one triangle and "applies" it to the other triangle. By this we understand that he means to pick up the first triangle, move it, and carefully place one vertex at a time on the corresponding vertices of the second triangle. This is often called Euclid's *method of superposition*. It is quite clear from an intuitive point of view that this operation should be possible, but again the objection can be raised that Euclid is using unstated assumptions about triangles. Over the years geometers have come to realize that the ability to move geometric objects around without distorting their shapes cannot be taken for granted. The need to include an explicit assumption about motions of triangles will be discussed further in Chapters 3 and 10.

Another interesting aspect of the proof is the fact that part of it is enclosed in square brackets. (See the words starting with, "For if . . . " in the third paragraph of the proof.) These words are in brackets because it is believed that they are not part of Euclid's original proof, but were inserted later.[2] They were added to justify Euclid's obvious assumption that there is only one straight line segment joining two points. Postulate I states that there exists a straight line joining two points, but here Euclid needs the stronger statement that there is exactly one such line. The fact that these words were added in antiquity is an indication that already then some readers of the *Elements* recognized that Euclid was using unstated assumptions.

Euclid's Proposition 16 is the result we now know as the Exterior Angle Theorem. This theorem and its proof will be discussed in Chapter 4. For now we merely point out that Euclid's proof depends on a relationship that appears to be obvious from the diagram provided, but which Euclid does not actually prove. Euclid wants to show that the interior angle $\angle BAC$ is smaller than the exterior angle $\angle ACD$. He first constructs the points E and F, and then uses the Vertical Angles Theorem (Proposition 15) and Side-Angle-Side to conclude that $\angle BAC$ is congruent to $\angle ACF$. Euclid assumes that F is in the interior of $\angle ACD$ and uses Common Notion 5 to conclude that $\angle BAC$ is smaller than $\angle ACD$. However, he provides no justification for the assertion that F is in the interior of angle $\angle ACD$. In Chapter 3 we carefully state postulates that will allow us to fill in this gap when we prove the Exterior Angle Theorem in Chapter 4.

---

2. See [Hea56, page 249].

1.6 A NEW VIEW OF THE FOUNDATIONS

The exhaustive study of Euclid's proofs described in the previous section forced mathematicians to realize that they had to be even more careful about the foundations of geometry than Euclid had been. In time this led to an entirely new understanding of the nature of postulates and their relationship to the world in which we live.

Euclid thought of his postulates as statements of self-evident truths about the real world. He stated some key geometric facts as postulates, but felt free to bring in other spatial relationships when they were needed and were obvious from the diagrams. In an effort to improve on Euclid, generations of geometers attempted to prove Euclid's Fifth Postulate using only Euclid's other postulates. Again and again they thought they had succeeded, only to find that there were hidden assumptions in their proofs. This forced them to acknowledge that they could only be sure of the status of their proofs if they stated *all* their assumptions and not just some of them. The goal became to develop a system of postulates that includes all the hypotheses needed to prove the propositions of geometry; reliance on any information that is not explicitly stated in the postulates could not be allowed.

As a result of that historical process, postulates were divorced from the real world, making them simply abstract logical assumptions. That is the point of view taken in this book and it is the way in which the foundations of mathematics are currently formulated. Euclid's basic logical structure was retained, but was strengthened and made more rigorous. The next chapter will set forth the modern perspective on the foundations.

1.7 SOME FINAL OBSERVATIONS ABOUT THE *ELEMENTS*

Before beginning our study of the modern formulation of the foundations of geometry, we make some additional observations about Euclid's *Elements*.

One aspect of Euclid's proofs that should be noted is the fact that each statement in the proof is justified by appeal to one of the postulates, common notions, definitions, or previous propositions. These references are placed immediately after the corresponding statements. They were probably not written explicitly in Euclid's original and therefore Heath encloses them in square brackets. This aspect of Euclid's proofs serves as an important model for the proofs we will write later in this course.

The words "Therefore etc." found near the end of the proofs are also not in Euclid's original. In the Greek view, the proof should culminate in a full statement of what had been proved. Thus Euclid's proof would have ended with a complete restatement of the conclusion of the proposition. Heath omits this reiteration of the conclusion and simply replaces it with "etc." Notice that the proof of Proposition 1 ends with the phrase "Being what it was required to do," while the proof of Proposition 4 ends with "Being what it was required to prove." The difference is that Proposition 1 is a construction while Proposition 4 is a logical implication. Later Heath uses the Latin abbreviations Q.E.F. and Q.E.D. for these phrases.

There are many features of Euclid's work that strike the modern reader as strange. One is the spare purity of Euclid's geometry. The points and lines are pure geometric forms that float in the plane with no fixed location. All of us have been trained since childhood to identify points on a line with numbers and points in the plane with pairs of numbers. That concept would have been foreign to Euclid; he did not mix the notions of number and point the way we do. The identification of number and point did not occur until the time of Descartes in the seventeenth century and it was not until the twentieth century that the real numbers were incorporated into the statements of the postulates of geometry. It is important to recognize this if we are to understand Euclid.

Euclid (really Heath) also uses language in a way that is different from contemporary usage. For example, what Euclid calls a line we would call a curve. We reserve the term *line* for what Euclid calls a "straight line." More precisely, what Euclid calls a straight line we would call a line segment (finitely long, with two endpoints). This distinction is more than just a matter of definitions; it indicates a philosophical difference. In Euclid, straight lines are potentially infinite in that they can always be extended to be as long as is needed for whatever construction is being considered, but he never considers the entire infinite line all at once. Since the time of Georg Cantor in the nineteenth century, mathematicians have been comfortable with sets that are actually infinite, so we usually think of the line as already being infinitely long and do not worry about the need to extend it.

Euclid chose to state his postulates in terms of straightedge and compass constructions. His propositions then often deal with the question of what can be constructed using those two instruments. For example, Proposition 1 really asserts the following: *Given a line segment, it is possible to construct, using only straightedge and compass, an equilateral triangle having the given segment as base.* In some ways Euclid identifies constructibility with existence. One of the major problems that the ancient Greeks never solved is the question of whether or not a general angle can be trisected. From a modern point of view the answer is obvious: any angle has a measure (in degrees, for example) which is a real number; simply dividing that real number by 3 gives us an angle that is one-third the original. But the question the Greeks were asking was whether the smaller angle can always be constructed from the original using only straightedge and compass. Such constructibility questions will be discussed in Chapter 9.

In this connection it is worthwhile to observe that the tools Euclid chose to use reflect the same pure simplicity that is evident throughout his work. His straightedge has no marks on it whatsoever. He did not allow a mark to be made on it that could be preserved when the straightedge is moved to some other location. In modern treatments of geometry we freely allow the use of a ruler, but we should be sure to note that a ruler is much more than a straightedge: It not only allows straight lines to be drawn, but it also measures distances at the same time. Euclid's compass, in the same way, is what we would now call a "collapsing" compass. It can be used to draw a circle with a given center and radius (where "radius" means a line segment with the center as one endpoint), but it cannot be moved to some other location and used to draw a different circle of the same radius. When the compass is picked up to be moved, it collapses and does not remember the radius of the previous circle. In contemporary treatments of geometry the compass has been supplemented by a protractor, which is a device for measuring angles. Euclid did not rely on numerical measurements of angles and he did not identify angles with the numbers that measure them the way we do.

EXERCISES

1.7.1 A *quadrilateral* is a four-sided figure in the plane. Consider a quadrilateral whose successive sides have lengths a, b, c, and d. Ancient Egyptian geometers used the formula

$$A = \frac{1}{4}(a + c)(b + d)$$

to calculate the area of a quadrilateral. Check that this formula gives the correct answer for rectangles but not for parallelograms.

1.7.2 An ancient Egyptian document, known as the *Rhind papyrus*, suggests that the area of a circle can be determined by finding the area of a square whose side has length $\frac{8}{9}$ the diameter of the circle. What value of π is implied by this formula? How close is it to the correct value?

1.7.3 The familiar Pythagorean Theorem states that if $\triangle ABC$ is a right triangle with right angle at vertex C and a, b, and c are the lengths of the sides opposite vertices A, B, and C, respectively, then $a^2 + b^2 = c^2$. Ancient proofs of the theorem were based on diagrams like those in Figure 1.6. Explain how the two diagrams together can be used to provide a proof for the theorem.

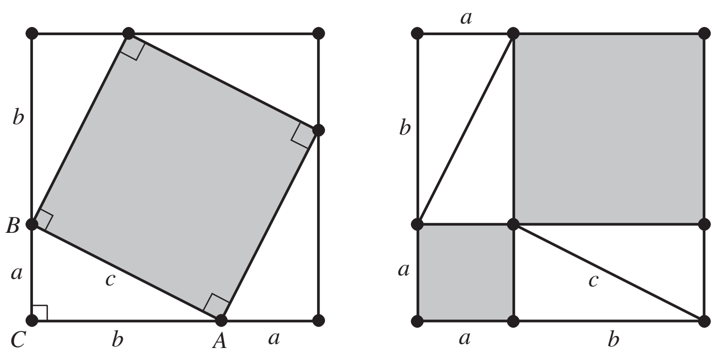


FIGURE 1.6 Proof of the Pythagorean Theorem

1.7.4 A *Pythagorean triple* is a triple (a, b, c) of positive integers such that $a^2 + b^2 = c^2$. A Pythagorean triple (a, b, c) is *primitive* if a, b, and c have no common factor. The tablet *Plimpton 322* indicates that the ancient Babylonians discovered the following method for generating all primitive Pythagorean triples. Start with relatively prime (i.e., no common factors) positive integers u and v, $u > v$, and then define $a = u^2 - v^2$, $b = 2uv$, and $c = u^2 + v^2$.

(a) Verify that (a, b, c) is a Pythagorean triple.

(b) Verify that a, b, and c are all even if u and v are both odd.

(c) Verify that (a, b, c) is a primitive Pythagorean triple in case one of u and v is even and the other is odd.

Every Pythagorean triple (a, b, c) with b even is generated by this Babylonian process. The proof of that fact is significantly more difficult than the exercises above but can be found in most modern number theory books.

1.7.5 The ancient Egyptians had a well-known interest in pyramids. According to the *Moscow papyrus*, they developed the following formula for the volume of a truncated pyramid with square base:

$$V = \frac{h}{3}(a^2 + ab + b^2).$$

In this formula, the base of the pyramid is an $a \times a$ square, the top is a $b \times b$ square, and the height of the truncated pyramid (measured perpendicular to the base) is h. One fact you learned in high school geometry is that the volume of a pyramid is one-third the area of the base times the height. Use that fact along with some high school geometry and algebra to verify that the Egyptian formula is exactly correct.

1.7.6 Explain how to complete the following constructions using only compass and straightedge. (You probably learned to do this in high school.)

(a) Given a line segment $\overline{AB}$, construct the perpendicular bisector of $\overline{AB}$.

(b) Given a line ℓ and a point P not on ℓ, construct a line through P that is perpendicular to ℓ.

(c) Given an angle $\angle BAC$, construct the angle bisector.

1.7.7 Can you prove the following assertions using only Euclid's postulates and common notions? Explain your answer.

(a) Every line has at least two points lying on it.

(b) For every line there is at least one point that does not lie on the line.

(c) For every pair of points $A \neq B$, there is only one line that passes through A and B.

1.7.8 Find the first of Euclid's proofs in which he makes use of his fifth postulate.

1.7.9 A *rhombus* is a quadrilateral in which all four sides have equal lengths. The *diagonals* are the line segments joining opposite corners. Use the first five propositions of Book I of the *Elements* to show that the diagonals of a rhombus divide the rhombus into four congruent triangles.

1.7.10 A *rectangle* is a quadrilateral in which all four angles have equal measures. (Hence they are all right angles.) Use the propositions in Book I of the *Elements* to show that the diagonals of a rectangle are congruent and bisect each other.

1.7.11 The following well-known argument illustrates the danger in relying too heavily on diagrams.[3] Find the flaw in the "Proof." (The proof uses familiar high school notation that will be explained later in this textbook. For example, $\overline{AB}$ denotes the segment from A to B and $\overleftrightarrow{AB}$ denotes the line through points A and B.)

False Proposition. If $\triangle ABC$ is any triangle, then side $\overline{AB}$ is congruent to side $\overline{AC}$.

Spurious Proof. Let ℓ be the line that bisects the angle $\angle BAC$ and let G be the point at which ℓ intersects $\overline{BC}$. Either ℓ is perpendicular to $\overline{BC}$ or it is not. We give a different argument for each case.

Assume, first, that ℓ is perpendicular to $\overline{BC}$ (Figure 1.7). Then $\triangle AGB \cong \triangle AGC$ by Angle-Side-Angle and therefore $\overline{AB} \cong \overline{AC}$.

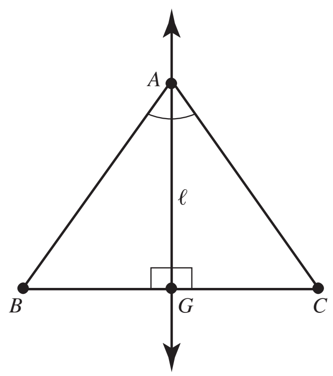


FIGURE 1.7 One possibility: the angle bisector is perpendicular to the base

Now suppose ℓ is not perpendicular to $\overline{BC}$. Let m be the perpendicular bisector of $\overline{BC}$ and let M be the midpoint of $\overline{BC}$. Then m is perpendicular to $\overleftrightarrow{BC}$ and ℓ is not, so m is

3. This fallacy is apparently due to W. W. Rouse Ball (1850–1925) and first appeared in the original 1892 edition of [BC87].

not equal to ℓ and m is not parallel to ℓ. Thus ℓ and m must intersect at a point D. Drop perpendiculars from D to the lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{AC}$ and call the feet of those perpendiculars E and F, respectively.

There are three possible locations for D: either D is inside $\triangle ABC$, D is on $\triangle ABC$, or D is outside $\triangle ABC$. The three possibilities are illustrated in Figure 1.8.

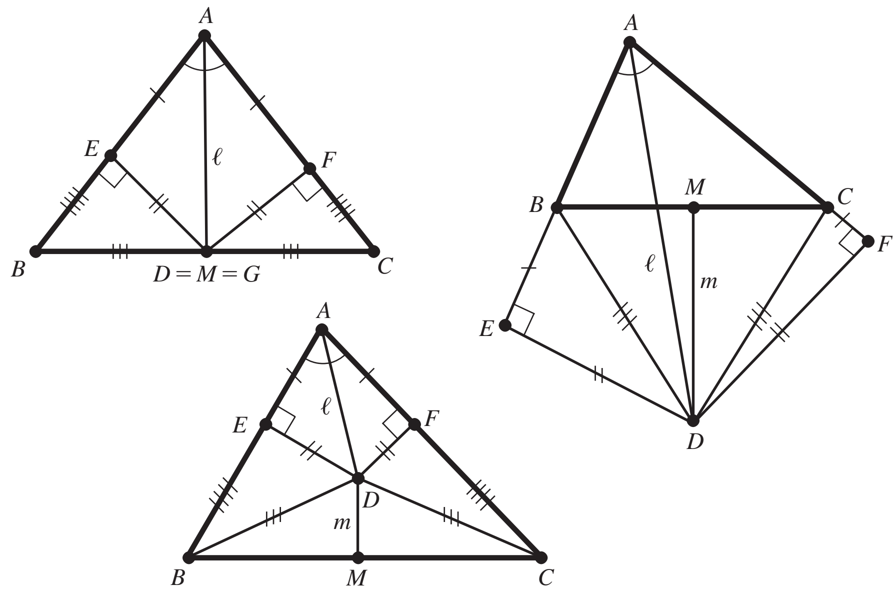


FIGURE 1.8 Three possible locations for D

The point D cannot lie on $\triangle ABC$. To see this, note that if D were on $\triangle ABC$, then it would be the case that $D = M = G$ because G is the only point at which ℓ intersects $\overleftrightarrow{BC}$ and M is the only point at which m intersects $\overleftrightarrow{BC}$. Now $\triangle ADE \cong \triangle ADF$ by Angle-Angle-Side, so $\overline{AE} \cong \overline{AF}$ and $\overline{DE} \cong \overline{DF}$. Also $\overline{BD} \cong \overline{CD}$ since $D = M$ is the midpoint of $\overline{BC}$. It follows from the Hypotenuse-Leg Theorem[4] that $\triangle BDE \cong \triangle CDF$ and therefore $\overline{BE} \cong \overline{CF}$. Hence $\overline{AB} \cong \overline{AC}$ by addition. But this means that $\triangle ADB \cong \triangle ADC$ by Side-Angle-Side and therefore $\angle ADB \cong \angle ADC$. Since $\angle ADB$ and $\angle ADC$ are supplementary angles, they must both be right angles. But that is impossible because ℓ is not perpendicular to $\overline{BC}$.

Now consider the case in which D is inside $\triangle ABC$. We have $\triangle ADE \cong \triangle ADF$ just as before, so again $\overline{AE} \cong \overline{AF}$ and $\overline{DE} \cong \overline{DF}$. Also $\triangle BMD \cong \triangle CMD$ by Side-Angle-Side and hence $\overline{BD} \cong \overline{CD}$. Applying the Hypotenuse-Leg Theorem again gives $\triangle BDE \cong \triangle CDF$ and therefore $\overline{BE} \cong \overline{CF}$ as before. It follows that $\overline{AB} \cong \overline{AC}$ by addition.

Finally consider the case in which D is outside $\triangle ABC$. Once again we have $\triangle ADE \cong \triangle ADF$ by Angle-Angle-Side, so again $\overline{AE} \cong \overline{AF}$ and $\overline{DE} \cong \overline{DF}$. Just as before, $\triangle BMD \cong \triangle CMD$ by Side-Angle-Side and hence $\overline{BD} \cong \overline{CD}$. Applying the Hypotenuse-Leg Theorem gives $\triangle BDE \cong \triangle CDF$ and therefore $\overline{BE} \cong \overline{CF}$. It then follows that $\overline{AB} \cong \overline{AC}$, this time by subtraction. ■

4. The Hypotenuse-Leg Theorem states that if the hypotenuse and leg of one right triangle are congruent to the corresponding parts of a second right triangle, then the triangles are congruent. It is a correct theorem; this is not the error in the proof.

SUGGESTED READING

- Chapters 1 and 2 of *Journey Through Genius*, [Dun91].
- Part I (pages 1–49) of *Euclid's Window*, [Mlo01].
- Chapters 1–4 of *Geometry: Our Cultural Heritage*, [Hol02].
- Chapters I– IV of *Mathematics in Western Culture*, [Kli53].
- Chapters 1 and 2 of *The Non-Euclidean Revolution*, [Tru01].
- Chapters 1 and 2 of *A History of Mathematics*, [Kat98].

This page intentionally left blank

CHAPTER 2

Axiomatic Systems and Incidence Geometry

Over the years that have passed since he wrote his *Elements*, Euclid's program for organizing geometry has been refined into what is called an *axiomatic system*. The basic structure of a modern axiomatic system was inspired by Euclid's method of organization, but there are significant ways in which an axiomatic system differs from Euclid's scheme.

This chapter examines the various parts of an axiomatic system and explains their relationships. Those relationships are illustrated by a fundamental example known as *incidence geometry*. By doing a preliminary study of axioms and relations among them in the simple, uncomplicated setting of incidence geometry we are better able to understand how an axiomatic system works and what it means to say that one axiom is independent of some others. This lays the groundwork for the presentation of plane geometry as an axiomatic system in Chapter 3 and also prepares the way for later chapters where it is proved that Euclid's Fifth Postulate is independent of his other postulates.

An important feature of the axiomatic method is proof. The chapter contains a review of some basic principles used in the construction of proofs and also sets out the distinctive style of written proof that will be used in this book. Incidence geometry provides a convenient setting in which to practice some of the proof-writing skills that will be required later.

2.1 THE STRUCTURE OF AN AXIOMATIC SYSTEM

An axiomatic system consists of undefined terms, definitions, axioms, theorems, and proofs. We will examine each of these parts separately.

Undefined and defined terms

The first part of an axiomatic system is a list of *undefined terms*. These are the technical words that will be used in the subject. Euclid attempted to define all his terms, but we now recognize that it is not possible to achieve the goal of defining every term. A standard dictionary appears to contain a definition of every word in a language, but there will inevitably be some circularity in the definitions because every definition uses words that must themselves be defined. Rather than attempting to define every term we will use, we simply take certain key words to be undefined and work from there.

In geometry, we usually take such words as *point* and *line* to be undefined. In other parts of mathematics, the words *set* and *element of* are often undefined. When the real numbers are treated axiomatically, the term *real number* itself is sometimes undefined.

Even though some words are left undefined, there is still a place for definitions and defined terms in an axiomatic system. The aim is to start with a minimal number of undefined terms and then to define other technical words using the original undefined terms and previously defined terms. One role of definitions is just to allow statements to be made concisely. For example, we will define three points to be collinear if there is one line such that all three points lie on that line. It is much more clear and concise to say that three points are noncollinear than it is to say that there does not exist a single line such that all three points lie on that line. Another function of definitions is to identify and highlight key structures and concepts.

Axioms

The second part of an axiomatic system is a list of *axioms*. The words *axiom* and *postulate* are used interchangeably and mean exactly the same thing in this book.[1] An axiom is a statement that is accepted without proof. Axioms are where the subject begins. Everything else should be logically deduced from them.

The axioms limit the way in which the undefined terms can be interpreted. Thus, for example, we do not define exactly what a point or a line is, but in the axioms for geometry we spell out those properties of points and lines that will be used in our development of geometry. In that limited sense the axioms serve to define the undefined terms.

All relevant assumptions are to be stated in the axioms and the only properties of the undefined terms that may be used in the subsequent development of the subject are those that are explicitly spelled out in the axioms. Hence we will allow ourselves to use those and only those properties of points and lines that have been stated in our axioms—any other properties or facts about points and lines that we know from our intuition or previous experience are not to be used until and unless they have been proven to follow from the axioms.

One of the goals of this course is to present plane geometry as an axiomatic system. This will require a much more extensive list of axioms than Euclid used. The reason for this is that we must include *all* the assumptions that will be needed in the proofs and not allow ourselves to rely on diagrams or any intuitive but unstated properties of points and lines the way Euclid did.

Theorems and proofs

The final part, usually by far the largest part, of an axiomatic system consists of the theorems and their proofs. Again there are two different words that are used synonymously: the words *theorem* and *proposition* will mean the same thing in this course.[2] In this third part of an axiomatic system we work out the logical consequences of the axioms.

---

1. Generally *postulate* will be used when a particular assumption is being stated or referred to by name, while the word *axiom* will be used in a more generic sense to refer to unproven assumptions.
2. There are also two other words that are used for theorem. A *lemma* is a theorem that is stated as a step toward some more important result. Usually a lemma is not an end in itself but is used as a way to organize a complicated proof by breaking it down into steps of manageable size. A *corollary* is a theorem that can be quickly and easily deduced from a previously stated theorem.

Just as in Euclid's *Elements*, there is a strict logical organization that applies. The first theorem is proved using only the axioms. The second theorem is proved using the first theorem together with the axioms, and so on.

Later in the chapter we will have much more to say about theorems and proofs as well as the rules of logic that are to be used in proofs.

Interpretations and models

In an axiomatic system the undefined terms do not in themselves have any definite meaning, except what is explicitly stated in the axioms. The terms may be interpreted in any way that is consistent with the axioms. An *interpretation* of an axiomatic system is a particular way of giving meaning to the undefined terms in that system. An interpretation is called a *model* for the axiomatic system if the axioms are correct (true) statements in that interpretation. Since the theorems in the system were all logically deduced from the axioms and nothing else, we know that all the theorems will automatically be correct and true statements in any model.

We say that a statement in our axiomatic system is *independent* of the axioms if it is not possible to either prove or disprove the statement as a logical consequence of the axioms. A good way to show that a statement is independent of the axioms is to exhibit one model for the system in which the statement is true and another model in which it is false. As we shall see, that is exactly the way in which it was eventually shown that Euclid's Fifth Postulate is independent of Euclid's other postulates.

The axioms in an axiomatic system are said to be *consistent* if no logical contradiction can be derived from them. This is obviously a property we would want our axioms to have. Again it is a property that can be verified using models. If there exists a model for an axiomatic system, then the system must be consistent. The existence of a model for Euclidean geometry and thus the consistency of Euclid's postulates was taken for granted until the nineteenth century. Our study of geometry will repeat the historical pattern: We will first study various geometries as axiomatic systems and only address the questions of consistency and existence of models later, in Chapter 11.

2.2 AN EXAMPLE: INCIDENCE GEOMETRY

In order to clarify what an axiomatic system is, we study the important example of *incidence geometry*. For now we simply look at the axioms and various models for this system; a more extensive discussion of theorems and proofs in incidence geometry is delayed until later in the chapter.

Let us take the three words *point, line*, and *lie on* (as in "point P lies on line ℓ") to be our undefined terms. The word *incident* is also used in place of *lie on*, so the two statements "P lies on ℓ" and "P is incident with ℓ" mean the same thing. For that reason the axioms for this relationship are called *incidence axioms*. One advantage of the word *incident* is that it can be used symmetrically: We can say that P is incident with ℓ or that ℓ is incident with P; both statements mean exactly the same thing.

There are three incidence axioms. When we say (in the axiom statements) that *P and Q are distinct points*, we simply mean that they are not the same point.

Incidence Axiom 1. For every pair of distinct points P and Q there exists exactly one line ℓ such that both P and Q lie on ℓ.

Incidence Axiom 2. For every line ℓ there exist at least two distinct points P and Q such that both P and Q lie on ℓ.

Incidence Axiom 3. There exist three points that do not all lie on any one line.

The axiomatic system with the three undefined terms and the three axioms listed above is called *incidence geometry*. We usually also call a model for the axiomatic system *an incidence geometry* and an interpretation of the undefined terms is called *a geometry*. Before giving examples of incidence geometries, it is convenient to introduce a defined term.

Definition 2.2.1. Three points A, B, and C are *collinear* if there exists one line ℓ such that all three of the points A, B, and C lie on ℓ. The points are *noncollinear* if there is no such line ℓ.

Using this definition we can give a more succinct statement of Incidence Axiom 3: *There exist three noncollinear points.*

EXAMPLE 2.2.2 ***Three-point geometry***

Interpret *point* to mean one of the three symbols A, B, C; interpret *line* to mean a set of two points; and interpret *lie on* to mean "is an element of." In this interpretation there are three lines, namely $\{A, B\}$, $\{A, C\}$, and $\{B, C\}$. Since any pair of distinct points determines exactly one line and no one line contains all three points, this is a model for incidence geometry. □

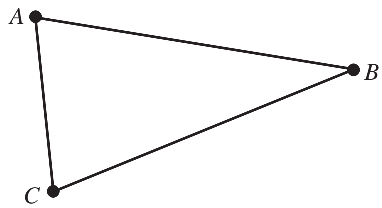


FIGURE 2.1 Three-point geometry

Be sure to notice that this "geometry" contains just three points. It is an example of a *finite geometry*, which is a geometry that contains only a finite number of points. It is customary to picture such geometries by drawing a diagram in which the points are represented by dots and the lines by segments joining them. So the diagram for the three-point plane looks like a triangle (see Figure 2.1). Don't be misled by the diagram: the "points" on the sides of the triangle are *not* points in the three-point plane. The diagram is strictly schematic, meant to illustrate relationships, and is not to be taken as a literal picture of the geometry.

EXAMPLE 2.2.3 ***The three-point line***

Interpret *point* to mean one of the three symbols A, B, C, but this time interpret *line* to mean the set of all points. This geometry contains only one line, namely $\{A, B, C\}$. In this interpretation Incidence Axioms 1 and 2 are satisfied, but Incidence Axiom 3 is not satisfied. Hence the three-point line is not a model for incidence geometry (Figure 2.2). □

EXAMPLE 2.2.4 ***Four-point geometry***

Interpret *point* to mean one of the four symbols A, B, C, D; interpret *line* to mean a set of two points and interpret *lie on* to mean "is an element of." In this interpretation there

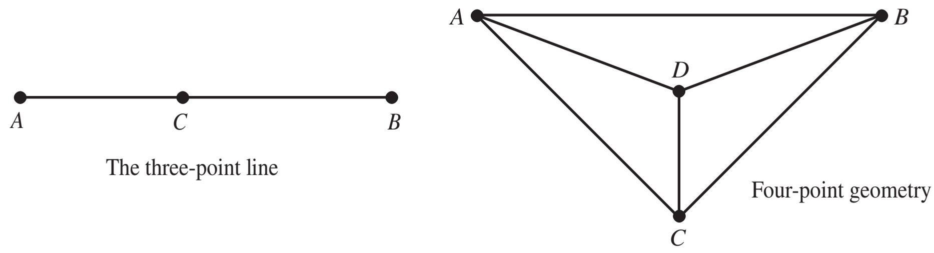


FIGURE 2.2 Two interpretations of the terms of incidence geometry

are six lines, namely $\{A, B\}$, $\{A, C\}$, $\{A, D\}$, $\{B, C\}$, $\{B, D\}$, and $\{C, D\}$. Since any pair of distinct points determines exactly one line and no one line contains three distinct points, this is a model for incidence geometry (Figure 2.2). □

EXAMPLE 2.2.5 ***Five-point geometry***

Interpret *point* to mean one of the five symbols A, B, C, D, E; interpret *line* to mean a set of two points and interpret *lie on* to mean "is an element of." In this interpretation there are ten lines. Again any pair of distinct points determines exactly one line and no one line contains three distinct points, so this is also a model for incidence geometry. □

We could continue to produce n-point geometries for increasingly large values of n, but we will stop with these three because they illustrate all the possibilities regarding parallelism that will be studied in the next section.

EXAMPLE 2.2.6 ***The interurban***

In this interpretation there are three *points*, namely the cities of Grand Rapids, Holland, and Muskegon (three cities in western Michigan). A *line* consists of a railroad line from one city to another. There is one railroad line joining each pair of distinct cities, for a total of three lines. Again, this is a model for incidence geometry. □

EXAMPLE 2.2.7 ***Fano's geometry***

Interpret *point* to mean one of the seven symbols A, B, C, D, E, F, G; interpret *line* to mean one of the seven three-point sets listed below and interpret *lie on* to mean "is an element of." The seven lines are

$$\{A, B, C\}, \{C, D, E\}, \{E, F, A\}, \{A, G, D\}, \{C, G, F\}, \{E, G, B\}, \{B, D, F\}.$$

All three incidence axioms hold in this interpretation, so Fano's geometry[3] is another model for incidence geometry. □

The illustration of Fano's geometry (Figure 2.3) shows one of the lines as curved while the others are straight. It should be recognized that this is an artifact of the schematic diagram we use to picture the geometry and is not a difference in the lines themselves. A line is simply a set of three points; the curves in the diagram are meant to show visually which points lie together on a line and are not meant to indicate anything about straightness. In fact, the word "straight" is not defined in incidence geometry and straightness is not a part of this geometry.

3. Named for Gino Fano, 1871–1952.

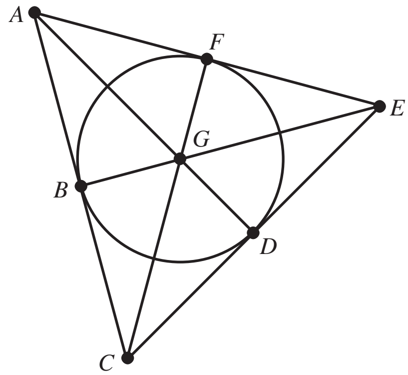


FIGURE 2.3 Fano's geometry

The examples of interpretations given so far illustrate the fact that the undefined terms in a given axiomatic system can be interpreted in widely different ways. No one of the models is preferred over any of the others. Notice that three-point geometry and the interurban are essentially the same; the names for the points and lines are different, but all the important relationships are the same. We could easily construct a correspondence from the set of points and lines of one model to the set of points and lines of the other model. The correspondence would preserve all the relationships that are important in the geometry (such as incidence). Models that are related in this way are called *isomorphic* models and a function between them that preserves all the geometric relationships is an *isomorphism*.

All the interpretations described so far have been finite geometries. Of course the geometries with which we are most familiar are not finite. We next describe three infinite geometries.

EXAMPLE 2.2.8 ***The Cartesian plane***

In this geometry a *point* is defined to be any ordered pair (x, y) of real numbers. A *line* is the collection of points whose coordinates satisfy a linear equation of the form $ax + by + c = 0$, where a, b, and c are real numbers and a and b are not both 0. More specifically, three real numbers a, b, and c, with a and b not both 0, determine the line ℓ consisting of all pairs (x, y) such that $ax + by + c = 0$; i.e.,

$$\ell = \{(x, y) \mid ax + by + c = 0\}.$$

A point (x, y) is said to *lie on* the line ℓ if the coordinates of the point satisfy the equation for ℓ. This is a model for incidence geometry; it is just the coordinate (or Cartesian) plane from high school Euclidean geometry. A complete verification that this interpretation satisfies the three incidence axioms requires some algebraic calculation (Exercise 2.4.8). We will use the symbol $\mathbb{R}^2$ to denote the set of points in the Cartesian plane (Figure 2.4). □

EXAMPLE 2.2.9 ***The sphere***

Interpret *point* to mean a point on the surface of a round 2-sphere in three-dimensional space. Specifically, a point is an ordered triple (x, y, z) of real numbers such that $x^2 + y^2 + z^2 = 1$. A *line* is interpreted to mean a great circle on the sphere and *lie on* is again

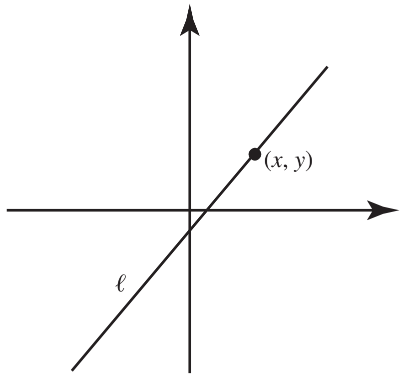


FIGURE 2.4 The Cartesian plane

interpreted to mean "is an element of." We will use the symbol $\mathbb{S}^2$ to denote the set of points on the sphere.

A *great circle* is a circle on the sphere whose radius is equal to that of the sphere. Alternatively, a great circle is the intersection of a plane through the origin in 3-space with the sphere. Two points on the sphere are *antipodal* (or opposite) if they are the two points at which a line through the origin intersects the sphere. Two given antipodal points on the sphere lie on an infinite number of different great circles; hence this geometry does not satisfy Incidence Axiom 1. If A and B are two points on the sphere that are not antipodal, then A and B determine a unique plane through the origin in 3-space and thus lie on a unique great circle. Hence "most" pairs of points determine a unique line in this geometry (Figure 2.5).

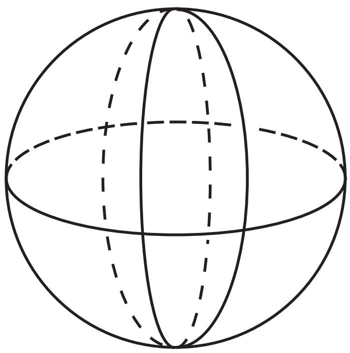


FIGURE 2.5 The sphere

Since this interpretation does not satisfy Incidence Axiom 1, it is not a model for incidence geometry. Note that Incidence Axioms 2 and 3 are correct statements in this interpretation. Another important observation about the sphere is that there are no parallel lines: any two distinct great circles on the sphere intersect in a pair of points. □

EXAMPLE 2.2.10 ***The Klein disk***

Interpret *point* to mean a point in the Cartesian plane that lies inside the unit circle. In other words, a point is an ordered pair (x, y) of real numbers such that $x^2 + y^2 < 1$. A *line*

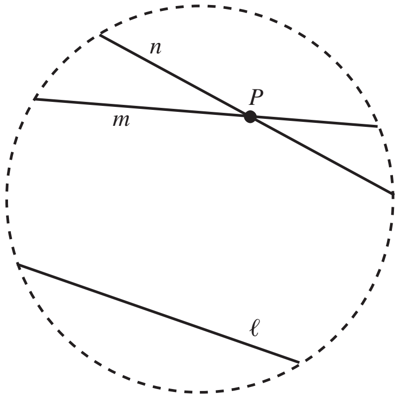


FIGURE 2.6 The Klein disk

is the part of a Euclidean line that lies inside the circle and *lie on* has its usual Euclidean meaning. This is a model for incidence geometry (Figure 2.6).

The Klein disk is an infinite model for incidence geometry, just like the familiar Cartesian plane is. (In this context *infinite* means that the number of points is unlimited, not that distances are unbounded.) The two models are obviously different in superficial ways. But they are also quite different with respect to some of the deeper relationships that are important in geometry. We illustrate this in the next section by studying parallel lines in each of the various geometries we have described. □

2.3 THE PARALLEL POSTULATES IN INCIDENCE GEOMETRY

Next we investigate parallelism in incidence geometry. The purpose of the investigation is to clarify what it means to say that the Euclidean Parallel Postulate is independent of the other axioms of geometry.

We begin with a definition of the word *parallel*, which becomes our second defined term in incidence geometry. In high school geometry parallel lines can be characterized in many different ways, so you may recall several definitions of *parallel*. In the context of everything that is assumed in high school geometry those definitions are logically equivalent and can be used interchangeably. But we have made no assumptions, other than those stated in the incidence axioms, so we must choose one of the definitions and make it the official definition of parallel. We choose the simplest characterization: the lines do not intersect. That definition fits best because it can be formulated using only the undefined terms of incidence geometry. It is also Euclid's definition of what it means for two lines in a plane to be parallel (Definition 23, Appendix A). Obviously this would not be the right definition to use for lines in 3-dimensional space, but the geometry studied in this book is restricted to the geometry of the two-dimensional plane. Note that, according to this definition of parallel, a line is not parallel to itself.

Definition 2.3.1. Two lines ℓ and m are said to be *parallel* if there is no point P such that P lies on both ℓ and m. The notation for parallelism is $\ell \parallel m$.

There are three different parallel postulates that will be useful in this course. The first is called the Euclidean Parallel Postulate, even though it is not actually one of Euclid's postulates. We will see later that (in the right context) it is logically equivalent to Euclid's Fifth Postulate. This formulation of the Euclidean Parallel Postulate is often called *Playfair's Postulate* (see page 105).

Euclidean Parallel Postulate. For every line ℓ and for every point P that does not lie on ℓ, there is exactly one line m such that P lies on m and $m \parallel \ell$.

There are other possibilities besides the Euclidean one. We state two of them.

Elliptic Parallel Postulate. For every line ℓ and for every point P that does not lie on ℓ, there is no line m such that P lies on m and $m \parallel \ell$.

Hyperbolic Parallel Postulate. For every line ℓ and for every point P that does not lie on ℓ, there are at least two lines m and n such that P lies on both m and n and both m and n are parallel to ℓ.

These are not new axioms for incidence geometry. Rather they are additional statements that may or may not be satisfied by a particular model for incidence geometry. We illustrate with several examples.

EXAMPLE 2.3.2 ***Parallelism in three-, four-, and five-point geometries***

In three-point geometry any two lines intersect. Therefore there are no parallel lines and this model satisfies the Elliptic Parallel Postulate.

Each line in four-point geometry is disjoint from exactly one other line. Thus, for example, the line $\{A, B\}$ is parallel to the line $\{C, D\}$ and no others. There is exactly one parallel line that is incident with each point that does not lie on $\{A, B\}$. Since the analogous statement is true for every line, four-point geometry satisfies the Euclidean Parallel Postulate.

Consider the line $\{A, B\}$ in five-point geometry and the point C that does not lie on $\{A, B\}$. Observe that C lies on two different lines, namely $\{C, D\}$ and $\{C, E\}$, that are both parallel to $\{A, B\}$. Since this happens for every line and for every point that does not lie on that line, five-point geometry satisfies the Hyperbolic Parallel Postulate. □

EXAMPLE 2.3.3 ***Parallelism in the Cartesian plane, sphere, and Klein disk***

The Euclidean Parallel Postulate is true in the Cartesian plane $\mathbb{R}^2$. The fact that the Cartesian plane satisfies the Euclidean Parallel Postulate is probably familiar to you from your high school geometry course. A proof of that statement would be an exercise in analytic geometry.

The sphere $\mathbb{S}^2$ satisfies the Elliptic Parallel Postulate. The reason for this is simply that there are no parallel lines on the sphere (any two great circles intersect).

The Klein disk satisfies the Hyperbolic Parallel Postulate. The fact that the Klein disk satisfies the Hyperbolic Parallel Postulate is illustrated in Figure 2.6. In that diagram, point P lies on both lines m and n and both m and n are parallel to ℓ. □

Conclusion

We can conclude from the preceding examples that each of the parallel postulates is independent of the axioms of incidence geometry. For example, the fact that there are some models for incidence geometry that satisfy the Euclidean Parallel Postulate and there are other models that do not shows that neither the Euclidean Parallel Postulate nor its negation can be proved as a theorem in incidence geometry. The examples make it clear that it would be fruitless to try to prove any of the three parallel postulates using only the axioms of incidence geometry.

This is exactly how it was eventually shown that Euclid's Fifth Postulate is independent of his other postulates. Understanding that proof is one of the major goals of this course. Later in the book we will construct two models for geometry, both of which satisfy all of Euclid's assumptions other than his fifth postulate. One of the models satisfies Euclid's Fifth Postulate while the other does not. (It satisfies the Hyperbolic Parallel Postulate.) This shows that it is impossible to prove Euclid's Fifth Postulate using only Euclid's other postulates and assumptions. It will take us most of the course to fully develop those models. One of the reasons it is such a difficult task is that we have to dig out *all* of Euclid's assumptions—not just the assumptions stated in his postulates, but all the unstated ones as well.

2.4 AXIOMATIC SYSTEMS AND THE REAL WORLD

An axiomatic system, as defined in this chapter, is obviously just a refinement of Euclid's system for organizing geometry. It should be recognized, however, that these refinements have profound implications for our understanding of the place of mathematics in the world.

The ancient Greeks revolutionized geometry by making it into an abstract discipline. Before that time, mathematics and geometry had been closely tied to the physical world. Geometry was the study of one aspect of the real world, just like physics or astronomy. In fact, the word *geometry* literally means "to measure the earth."

Later Greek geometry, on the other hand, is about relationships between ideal, abstract objects. In this view, geometry is not just about the physical world in which we live our everyday lives, but it also gives us information about an ideal world of pure forms. In the view of Greek philosophers such as Plato, this ideal world was, if anything, more real than the physical world of our existence. The relationships in the ideal world are eternal and pure. The Greeks presumably thought of a postulate as a statement about relationships that really pertain in that ideal world. The postulates are true statements that can be accepted without proof because they are self-evident truths about the way things really are in the ideal world. So the Greeks distanced geometry from the physical world by making it abstract, but at the same time they kept it firmly rooted in the real world where it could give them true and reliable information about actual spatial relationships.

We have no direct knowledge of how Euclid himself understood the significance of his geometry. All we know about his thinking is what we find written in the *Elements* and those books are remarkably terse by modern standards. But Euclid lived approximately 100 years after Plato, so it seems reasonable to assume that he was influenced by Plato's ideas. In any event, it is quite clear from reading the *Elements* that Euclid thought of geometry as being about real things and that is precisely why he felt free to use intuitively obvious facts about points and lines in his proofs even though he had not stated these facts as axioms.

The view of geometry as an axiomatic system (as described in this chapter) moves us well beyond the Platonic view. In our effort to spell out completely what our assumptions are, we have been led to make geometry much more relative and detached from reality. We do not apply the terms true or false to the axioms in any absolute sense. An axiom is simply a statement that may be true or may be false in any particular situation, it just depends on how we choose to interpret the undefined terms. Thus our efforts to introduce abstraction and rigor into geometry have led us to drain the meaning out of such everyday terms as point and line. Since the words can now mean just about anything we want them to, we must wonder whether they any longer have any real content.

The naïve view is that geometry is the study of space and spatial relationships. We usually think of geometry as a science that gives us true and reliable information about the world in which we live. The view of geometry as an axiomatic system detached from the real world is a bit disturbing to most of us.

Some mathematicians have promoted the view that mathematics is just a logical game in which we choose an arbitrary set of axioms and then see what we can deduce using the rules of logic. Most professional mathematicians, however, have a profound sense that the mathematics they study is about real things. The fact that mathematics has such incredibly powerful and practical applications is evidence that it is much more than a game.

It is surprisingly difficult to resolve the kinds of philosophical issues that are raised by these observations. The mathematical community's thinking on these matters has evolved over time and there have been several amazing revolutions in the conventional understanding of what the correct views should be. Those views will be explored in this book as the historical development of geometry unfolds. We do not attempt to give definitive answers; instead we simply raise the questions and encourage the reader to think about them. In particular, the following questions should be recognized and should be kept in mind as the development of geometry is worked out in this book.

1. Are the theorems of geometry true statements about the world in which we live?
2. What physical interpretation should we attach to the terms *point* and *line*?
3. What are the axioms that describe the geometry of the space in which we live?
4. Is it worthwhile to study arbitrary axiom systems or should we restrict our attention to just those axiom systems that appear to describe the real world?

At this point you might be asking yourself why it would be thought desirable to make mathematics so abstract and therefore to get into the kind of difficult issues that have been raised here. That is one question we can answer. The answer is that abstraction is precisely what gives mathematics its power. By identifying certain key features in a given situation, listing exactly what it is about those features that is to be studied, and then studying them in an abstract setting detached from the original context, we are able to see that the same kinds of relationships hold in many apparently different contexts. We are able to study the important relationships in the abstract without a lot of other irrelevant information cluttering up the picture and obscuring the underlying structure. Once things have been clarified in this way, the kind of logical reasoning that characterizes mathematics becomes an incredibly powerful and effective tool. The history of mathematics is full of examples of surprising practical applications of mathematical ideas that were originally discovered and developed by people who were completely unaware of the eventual applications.

EXERCISES

2.4.1 It is said that Hilbert once illustrated his contention that the undefined terms in a geometry should not have any inherent meaning by claiming that it should be possible to replace *point* by *beer mug* and *line* by *table* in the statements of the axioms. Consider three friends sitting around one table. Each person has one beer mug. At the moment all the beer mugs are resting on the table. Suppose we interpret point to mean beer mug, line to mean the table, and lie on to mean resting on. Is this a model for incidence geometry? Explain. Is this interpretation isomorphic to any of the examples in the text?

2.4.2 One-point geometry contains just one point and no line. Which incidence axioms does one-point geometry satisfy? Explain. Which parallel postulates does one-point geometry satisfy? Explain.

2.4.3 Two-point geometry consists of two points and one line. Both points lie on that line. Which incidence axioms does two-point geometry satisfy? Explain. Which parallel postulates does two-point geometry satisfy? Explain.

2.4.4 Consider a small mathematics department consisting of Professors Alexa, Bailey, Curtis, and Duarte with three committees: curriculum committee, personnel committee, and social committee. Interpret *point* to mean a member of the department, interpret *line* to be a departmental committee, and interpret *lie on* to mean that the faculty member is a member of the specified committee.

(a) Suppose the committee memberships are as follows: Alexa, Bailey, and Curtis are on the curriculum committee; Alexa and Duarte are on the personnel committee; and Bailey and Curtis are on the social committee. Is this a model for incidence geometry? Explain.

(b) Suppose the committee memberships are as follows: Alexa, Bailey, and Curtis are on the curriculum committee; Alexa and Duarte are on the personnel committee; and Bailey and Duarte are on the social committee. Is this a model for incidence geometry? Explain.

(c) Suppose the committees are the same as in part (b) but a fourth committee, the promotion committee, is added. Curtis and Duarte are the members of the promotion committee. Is this a model for incidence geometry? Explain.

(d) Suppose the committee memberships are as follows: Alexa and Bailey are on the curriculum committee, Alexa and Curtis are on the personnel committee, Duarte and Curtis are on the social committee, and Bailey and Duarte are on the promotions committee. Is this a model for incidence geometry? Explain.

2.4.5 A three-point geometry is an incidence geometry that satisfies the following additional axiom: *There exist exactly three points.*

(a) Find a model for three-point geometry.

(b) How many lines does any model for three-point geometry contain? Explain.

(c) Explain why any two models for three-point geometry must be isomorphic. (An axiomatic system with this property is said to be *categorical*.)

2.4.6 Interpret *point* to mean one of the four vertices of a square, *line* to mean one of the sides of the square, and *lie on* to mean that the vertex is an endpoint of the side. Which incidence axioms hold in this interpretation? Which parallel postulates hold in this interpretation?

2.4.7 Draw a schematic diagram of five-point geometry (see Example 2.2.5).

2.4.8 Verify that the Cartesian plane satisfies all three incidence axioms.

2.4.9 Which parallel postulate does Fano's geometry satisfy? Explain.

2.4.10 Which parallel postulate does the three-point line satisfy? Explain.

2.4.11 Must every incidence geometry satisfy at least one of the parallel postulates? Either explain why the answer is "yes" or give an example to show that the answer is "no."

2.4.12 Could an incidence geometry satisfy more than one of the parallel postulates? Explain.

2.4.13 Consider a finite model for incidence geometry that satisfies the following additional axiom: *Every line has exactly three points lying on it.* What is the minimum number of points in such a geometry? Explain your reasoning.

2.4.14 Find a finite model for incidence geometry in which there is one line that has exactly three points lying on it and there are other lines that have exactly two points lying on them.

2.4.15 Find interpretations for the words *point*, *line*, and *lie on* that satisfy the following conditions.

(a) Incidence Axioms 1 and 2 hold, but Incidence Axiom 3 does not.

(b) Incidence Axioms 2 and 3 hold, but Incidence Axiom 1 does not.

(c) Incidence Axioms 1 and 3 hold, but Incidence Axiom 2 does not.

2.4.16 For any interpretation of incidence geometry there is a *dual* interpretation. For each point in the original interpretation there is a line in the dual and for each line in the original there is a point in the dual. A point and line in the dual are considered to be incident if the corresponding line and point are incident in the original interpretation.

(a) What is the dual of the three-point plane? Is it a model for incidence geometry?

(b) What is the dual of the three-point line? Is it a model for incidence geometry?

(c) What is the dual of four-point geometry? Is it a model for incidence geometry?

(d) What is the dual of Fano's geometry?

2.5 THEOREMS, PROOFS, AND LOGIC

We now take a more careful look at the third part of an axiomatic system: the theorems and proofs. Both theorems and proofs require extra care. Most of us have enough experience with mathematics to know that the ability to write good proofs is a skill that must be learned, but we often overlook the fact that a necessary prerequisite to good proof-writing is good statements of theorems.

A major goal of this course is to teach the art of writing proofs and it is not expected that the reader is already proficient at it. The main way in which one learns to write proofs is by actually writing them, so the remainder of the book will provide lots of opportunities for practice. This section simply lays out a few basic principles and then those principles will be put to work in the rest of the course. The brief introduction provided in this section will not make you an instant expert at writing proofs, but it will equip you with the basic tools you need to get started. You should refer back to this section as necessary in the remainder of the course.

Mathematical language

An essential step on the way to the proof of a theorem is a careful statement of the theorem in clear, precise, and unambiguous language. To illustrate this point, consider the following proposition in incidence geometry.

Proposition 2.5.1. Lines that are not parallel intersect in one point.

Compare that statement with the following.

Restatement. If ℓ and m are two distinct lines that are not parallel, then there exists exactly one point P such that P lies on both ℓ and m.

Both are correct statements of the same theorem. But the second statement is much better, at least as far as we are concerned, because it clearly states where the proof should begin (with two lines ℓ and m such that $\ell \neq m$ and $\ell \nparallel m$) and where it should end (with a point P that lies on both ℓ and m). This provides the framework within which we can build a proof. Contrast that with the first statement. In the first statement it is much less clear how to begin a proof. In fact it is not possible to construct a proof until we have at least mentally translated the first statement into language that is closer to that in the second statement. When we start to do that we realize that the first statement is not precise enough. For example, it does not clearly say whether it is an assertion about two (or more?) particular lines or whether it is making an assertion that applies to all lines.

Writing good proofs requires clear thinking and clear thinking begins with careful statements. As a result we begin by examining the language used in the statements of theorems and only turn our attention to proof-writing after that.

Statements

In mathematics, the word *statement* refers to any assertion that can be classified as either true or false (but not both). Here is an example: Dan is tiny. The statements of geometry often involve assertions that objects (such as points or lines) satisfy certain conditions (such as parallelism). Such statements must be preceded by definitions of the terms used. For example, it is not possible to determine if the assertion *Dan is tiny* is true or false unless we have a precise definition of what *tiny* means in this context. It is obvious that the word *tiny* might mean one thing in one situation (for example, in microbiology) and something completely different in another context (such as astronomy). So we would need a definition of the form, "A person x is said to be *tiny* if . . . ." Once you have that definition, you can check to see whether or not a particular person named Dan satisfies the conditions in the definition.

Simple statements can be combined into compound statements using the words *and* and *or*. The use of *and* is easy to understand; it means that both the statements are true. The use of *or* in mathematics differs somewhat from the way the word is used in ordinary language. In mathematics *or* is always used in a nonexclusive way; it means that one or the other of the statements is true and allows the possibility that both are true. Consider the statement *Joan is old or Joan is rich*. As used in ordinary everyday English, this statement allows the possibility that Joan is both old and rich. Contrast that with the following statement: *Either you are for me or you are against me*. We understand from the tone of the statement that it means you are one or the other but not both. Thus the word *or* is ambiguous in ordinary English and its exact meaning depends on the context. Mathematical language eliminates such ambiguities and the word *or* always has the nonexclusive meaning when it is used in a mathematical statement.

We will often want to negate statements. Specifically, given one statement we will want to write down a second statement that asserts the opposite of the first. There is a sense in which it is easy to negate a statement: simply say, "It is not the case that . . . ." But this is not helpful. It is much more useful to observe that negation interchanges *and* and *or*. In other words, we have the following laws (for any statements S and T).

$$\text{not } (S \text{ and } T) = (\text{not } S) \text{ or } (\text{not } T)$$
$$\text{not } (S \text{ or } T) = (\text{not } S) \text{ and } (\text{not } T)$$

For example, if it is not true that $x > 0$ and $x < 1$, then either $x \leq 0$ or $x \geq 1$. The two rules stated above are known as *De Morgan's Laws*.

Propositional functions

An assertion such as "$x > 0$" does not, by itself, qualify as a statement in the technical sense defined above because it is neither true nor false until a value has been assigned to x. The assertion is really a function whose domain consists of numbers x and whose range consists of statements. For example, if we take $x = 1$, the result is the statement $1 > 0$ (which is true). On the other hand, taking $x = -1$ yields the statement $-1 > 0$ (which is false). Such a function is called a *propositional function*.[4] The function just described can be defined by the simple formula

$$P(x) = (x > 0).$$

Propositional functions can have more than one independent variable, as in $Q(x, y) = (x > y)$. In this example, $Q(1, 1)$ is a false statement while $Q(1, 0)$ is a true statement. This is precisely the kind of assertion we will encounter in the geometry course. For example, the assertion $\ell \parallel m$ should be interpreted as a propositional function whose domain is the set of ordered pairs of lines.

Quantifiers

One of the distinctions that must be made clear is whether you are asserting that *every* object of a certain type satisfies a condition or whether you are simply asserting that there is one that does. This is specified through the use of *quantifiers*. There are two quantifiers: the existential quantifier (written $\exists$) and the universal quantifier (written $\forall$).

A propositional function can be combined with a quantifier to yield a statement that has a truth value. For example, while $x > 0$ is neither true nor false, the statement $\exists x\ (x > 0)$ (read "there exists an x such that x is greater than zero") is true. On the other hand, $\forall x\ (x > 0)$ (read "for every x, x is greater than zero") is false.

Strictly speaking, we should specify the domain of discourse when using a quantifier. But it is common practice to omit mention of the domain when it can be inferred from the context. When the domain is not clear from the context, it should be explicitly specified; for example, $\exists x \in \mathbb{R}\ (x > 0)$.

Negation interchanges the two quantifiers. For example, consider this statement: Every angle is acute. Stated more precisely, it says, for every angle α, α is acute. The negation is this: There exists an angle α such that α is not acute. Here is another example: There exists a point that does not lie on ℓ. The negation is this statement: Every point lies on ℓ. Better yet is this statement: For every point P, P lies on ℓ. The rules for negating quantified statements are new versions of De Morgan's Laws:

$$\text{not } (\forall x\, P(x)) = \exists x (\text{not } P(x))$$
$$\text{not } (\exists x\, P(x)) = \forall x (\text{not } P(x))$$

Conditional statements

A *conditional statement* is a compound statement of the form "If . . . , then . . ." in which the first set of dots represents a statement called the *hypothesis* (or *antecedent*) and the

4. In some contexts the word "proposition" is used for what we are calling a "statement." That is the origin of the terminology *propositional* function. In this book the word "proposition" is almost always used as a synonym for "theorem."

second set of dots represents a statement called the *conclusion* (or *consequent*). The conditional statement "if H, then C" can also be formulated in words as "H implies C" or in symbols as $H \Rightarrow C$.

In a conditional statement, the hypothesis is always assumed to be universally quantified unless otherwise specified. In other words, if H and C are propositional functions, the statement $H(x) \Rightarrow C(x)$ means that for every x such that $H(x)$ is true, $C(x)$ is true as well. A simple example from high school algebra is the conditional statement, *If $x < 1$, then $x < 2$.* This statement is true because any value of x for which the propositional function $x < 1$ is true also makes the propositional function $x < 2$ true.

EXAMPLE 2.5.2 ***The Euclidean Parallel Postulate as a conditional statement***

The Euclidean Parallel Postulate can be reformulated as a conditional statement: *If ℓ is a line and P is a point that does not lie on ℓ, then there exists exactly one line m such that P lies on m and m is parallel to ℓ.* □

The statement $H \Rightarrow C$ really just rules out the possibility that H is true while C is false. Consider this statement: *If x is a real number and $x^2 < 0$, then $x = 3$.* It is considered to be true because the conclusion is true of every x for which the hypothesis is true. (There is no x for which the hypothesis is true.) In a case like this we usually say that the statement is *vacuously true* since the conditional statement is true only because there is no way the hypotheses can be true.

The negation of a conditional statement

Negating a conditional statement requires clear thinking. The statement $P \Rightarrow Q$ means that Q is true whenever P is. The negation of $P \Rightarrow Q$ is the assertion that it is possible for P to be true while Q is false. Be sure to note that the negation of a conditional statement is *not* another conditional statement. Consider the following example: *If x is irrational, then x^2 is irrational.* This statement is false[5] because there are some irrational numbers whose square is rational [e.g., $\sqrt{2}$ is irrational while $(\sqrt{2})^2$ is rational]. It is true that there are some irrational numbers whose squares are irrational, but it takes only one example to show that the conditional statement is false. For this reason we normally demonstrate that a conditional statement is false by producing a counterexample.

EXAMPLE 2.5.3 ***The negation of the Euclidean Parallel Postulate***

The following statement is the negation of the Euclidean Parallel Postulate: *There exists a line ℓ and there exists a point P that does not lie on ℓ such that there is not exactly one line m for which P lies on m and $m \parallel \ell$.* The statement "there is not exactly one line m for which P lies on m and $m \parallel \ell$" means that either there is no line with these properties or there are at least two lines with those properties. □

Converse and contrapositive

For every conditional statement there are two related statements called the *converse* and the *contrapositive*. The converse of $P \Rightarrow Q$ is $Q \Rightarrow P$ and the contrapositive is *not* $Q \Rightarrow$ *not* P. The converse to a conditional statement is an entirely different statement; the fact that $P \Rightarrow Q$ is true tells us nothing about whether or not the converse is true. On the other hand, the contrapositive is logically equivalent to the original statement. Here is a simple example: If $x = 2$, then $x^2 = 4$. This is a correct statement. Its converse, however,

5. The real theorem is this: *If x is rational, then x^2 is rational.*

is not correct. ($x^2 = 4 \not\Rightarrow x = 2$.) The contrapositive is this: If $x^2 \neq 4$, then $x \neq 2$. The contrapositive is a correct statement and is just a negative way of restating the original conditional statement.

Consider another simple example: If $x = 0$, then $x^2 = 0$. This time both the statement and its converse are true. Such statements are called *biconditional* statements and the phrase "if and only if" (abbreviated iff) is used to indicate that the implication goes both ways. In other words, P iff Q (or $P \Leftrightarrow Q$) means $P \Rightarrow Q$ and $Q \Rightarrow P$. Thus we could say that $x = 0$ iff $x^2 = 0$. An if-and-only-if statement is really two theorems in one and the proof should reflect this; there should normally be separate proofs of each of the two implications.

Truth tables

There is a sense in which the fact that a conditional statement and its contrapositive are equivalent is obvious, but it can be confusing to explain the equivalence because negations are piled on negations. A simple device that can be used to explain the equivalence is a *truth table*. This is a good way to think of it because the truth table also sheds light on the meaning of the conditional statement itself.

Consider the statement $H \Rightarrow C$. The hypothesis and the conclusion can each be either true or false. Thus there are four possibilities for H and C and the statement $H \Rightarrow C$ is true in three of the four cases. The various possibilities are shown in the following truth table.

H	C	$H \Rightarrow C$
True	True	True
True	False	False
False	True	True
False	False	True

It is the second half of the table that often confuses beginners; these are the cases in which the theorem is vacuously true. Since the conditional statement is true in three of the four cases, a proof is simply an argument that rules out the fourth possibility. If we now expand the table to include the negations of H and C as well as the contrapositive of the theorem we see that the contrapositive is true exactly when the theorem is. (The third and sixth columns are identical.)

H	C	$H \Rightarrow C$	not H	not C	not $C \Rightarrow$ not H
True	True	True	False	False	True
True	False	False	False	True	False
False	True	True	True	False	True
False	False	True	True	True	True

Sometimes it is more convenient to prove the contrapositive of a theorem than it is to prove the theorem itself. This is perfectly legitimate because the original statement is logically equivalent to the contrapositive.

Uniqueness

The word *unique* is often used in connection with the existential quantifier. For example, here is a statement that is important in geometry: For every line ℓ and for every point P there is a unique line m such that P lies on m and m is perpendicular to ℓ. The word *unique* in this statement indicates that there is exactly one line m satisfying the stated conditions. A proof of this assertion should have two parts. First, there should be a proof that there is a line m satisfying the conditions and, second, a proof that there cannot be two different lines m and n satisfying the conditions. The usual strategy for the second half of the proof is to start with the assumption that m and n are lines that satisfy the property and then to prove that m and n must, in fact, be equal to each other. The symbol ! is used to indicate uniqueness; so the notation "$\exists\,!\ldots$" should be read "there exists a unique . . . ."

Theorems

A *theorem* is a conditional statement that has been proved true. A conditional statement may be either true or false, but it is not called a theorem unless it is true and has been (or can be) proved. This means that there are no false theorems, just statements that have the form of theorems but turn out not to be theorems. Here is a good rule to follow: *Every theorem should be stated in if-then form.*[6]

A theorem does not assert that the conclusion is true without the hypothesis, only that the conclusion is true if the hypothesis is.

Proofs

A proof consists of a sequence of steps that lead us logically from the hypothesis to the conclusion. Each step should be justified by a reason. There are six kinds of reasons that can be given:

- by hypothesis
- by axiom
- by previous theorem
- by definition
- by an earlier step in this proof
- by one of the rules of logic

In this course we plan to assume much of what you already know about the algebra of real numbers. Thus the first several kinds of reasons listed above will often be stated as "by properties of real numbers" or "by algebra."

At the beginning of the course we will follow Euclid's practice of writing the reasons in parentheses after the statements. We will eventually drop that style as we develop more proficiency at writing and reading proofs. But for now, it is important to spell out all your justifications.

Writing proofs

In high school you may have learned to write your proofs in two columns, with the statements in one column and the reasons in another. We will not do that in this course, not even at the outset, because we are aiming to write proofs that can be read by fellow

6. As is the case with most rules, this one allows some exceptions. Here is a well-known theorem from calculus: *π is irrational*. In this case the hypotheses are hidden in the definition of π. While such a statement does qualify as a theorem, it is not a model we should adopt for this course. The practice of stating the hypotheses explicitly will serve us well as we learn to write proofs. The example does illustrate the fact that theorem statements are context dependent and there are often unstated hypotheses that are assumed in a given setting.

humans; in order to facilitate this, the proofs should be written in ordinary paragraph form. For the same reason we will not follow the high school custom of numbering the statements in a proof.

It is helpful to distinguish between the proof itself and the written argument that is used to communicate the proof to other people. The proof is a sequence of logical steps that lead from the hypothesis to the conclusion. The written argument lays out those steps for the reader, and the writer has an obligation to write them in a way that the reader can understand without undue effort. So the written proof is both a listing of the logical steps and an explanation of the reasoning that went into them.

Obviously you need to know who your audience is. You should assume that the reader is someone, like a fellow student, who has approximately the same background you have. It is important to remember that written proofs have a subjective aspect to them. They are written for a particular audience and how many details you include will depend on who is to read the proof. As you and the rest of the class learn more geometry together, you will share a larger and larger set of common experiences. You can draw on those experiences and assume that your readers will know many of the justifications for steps in the proofs. Later in the course you will be able to leave many of the reasons unstated; this will allow you to concentrate on the essential new ideas in a proof and not obscure them with a lot of detailed information that is already well known to your readers. But don't be too quick to jump ahead to that level. Our aim in this course is to lay out all our assumptions in the axioms and then to base our proofs on those assumptions and no others. Only by being explicit about our reasons for each step can we discipline ourselves to use only those assumptions and not bring in any hidden assumptions that are based on previous experience or on diagrams.

You are encouraged to include in your written proofs more than just a list of the logical steps in the proof. In order to make the proofs more readable, you should also include information for the reader about the structure of the logical argument you are using, what you are assuming, and where the proof is going. Such statements are not strictly necessary from a logical point of view, but they make an enormous difference in the readability of a proof. For now our goal is not so much to prove the theorems as it is to learn to write good proofs, so we will write more than is strictly necessary and not worry about the risk of being pedantic.

The beginning of a proof is marked by the word **Proof**. It is also a good idea to include an indication that you have reached the end of a proof. Traditionally the end of a proof was indicated with the abbreviation QED, which stands for the Latin phrase *quod erat demonstrandum* (which was to be demonstrated). In this book we mark the end of our proofs with the symbol ■.

Like each individual proof, the overall structure of the collection of theorems and proofs in an axiomatic system should be logical and sequential. Within any given proof, it is legitimate to appeal only to the axioms, a theorem that has been previously stated and proved, a definition that has been stated earlier, or to an earlier step in the same proof. The rules of logic that are listed as one possible type of justification for a step in a proof are the rules that are explained in this section. They include such rules as the rules for negating compound statements that were described above and the rules for indirect proofs that will be described below.

There is one last point related to the justification of the steps in a proof that is specific to this course in the foundations of geometry. We intend to build geometry on the real number system. Hence we will base many steps in our proofs on known facts about the real numbers. For example, if we have proved that $x + z = y + z$, we will want to conclude that $x = y$. Technically, this falls under the heading "by previous theorem," but we will usually say something like "by algebra" when we bring in some fact about the algebra of real numbers. Appendix E lists many of the important properties of the real numbers that are assumed in this course. A few of them have names (such as Trichotomy and the Archimedean Property) and those names should be mentioned when the properties are used.

Indirect proof

Indirect proof is one proof form that should be singled out for special consideration because you will find that it is one you will often want to use. The straightforward strategy for proving $P \Rightarrow Q$, called *direct proof*, is to start by assuming that P is true and then to use a series of logical deductions to conclude Q. But the statement $P \Rightarrow Q$ means that Q is true if P is, so the real purpose of the proof is to rule out the possibility that P is true while Q is false. The indirect strategy is to begin by assuming that P is true *and* that Q is false, and then to show that this leads to a logical contradiction. If it does, then we know that it is impossible for both P and the negation of Q to hold simultaneously and therefore Q must be true whenever P is. This indirect form of proof is called *proof by contradiction*. It goes by the official name *reductio ad absurdum*, which we will abbreviate as RAA.

The reason this proof form often works so well is that you have more information with which to work. In a direct argument you assume only the hypothesis P and work from there. In an indirect argument you begin by assuming both the original hypothesis P and also the additional hypothesis *not* Q. You can make use of both assumptions in your proof. In order to help clarify what is going on in an indirect proof, we will give a special name to the additional hypothesis *not* Q; we will call it the *RAA hypothesis* to distinguish it from the standard hypothesis P.

Indirect proofs are often confused with direct proofs of the contrapositive. They are not the same, however, since in a direct proof of the contrapositive we assume only the negation of the conclusion while in an indirect proof we assume and use both the hypothesis and the negation of the conclusion. Suppose we want to prove the theorem P implies Q. A direct proof of the contrapositive would start with *not* Q and conclude *not* P. One way to formulate the argument would be to start by assuming P (the hypothesis) and *not* Q (the RAA hypothesis), then to use the same proof as before to conclude *not* P and finally to end by saying that we must reject the RAA hypothesis because we now have both P and *not* P, an obvious contradiction. While this is logically correct, it is considered to be bad form and sloppy thinking; this way of organizing a proof should, therefore, usually be avoided.[7]

An even worse misuse of proof by contradiction is the following. Suppose again that we want to prove the theorem P implies Q. We assume the hypothesis P. Then we also suppose *not* Q (RAA hypothesis). After that we proceed to prove that P implies Q. At that point in the proof we have both Q and *not* Q. That is a contradiction, so we reject the RAA hypothesis and conclude Q. In this case the structure of an indirect argument has been erected around a direct proof, thus obscuring the real proof. Again this is sloppy thinking. It is an abuse of indirect proof and should *always* be avoided.

Use RAA proofs when they are helpful, but don't misuse them. A proof is often discovered as an indirect proof because we can suppose the conclusion is false and explore

7. We will occasionally find good reason to formulate our proofs this way even though it is generally not good practice.

the consequences. Once you have found a proof, you should reexamine it to see if the logic can be simplified and the essence of the proof presented more directly.

EXERCISES

2.5.1 Identify the hypothesis and conclusion of each of the following statements.

(a) If it rains, then I get wet.

(b) If the sun shines, then we go hiking and biking.

(c) If $x > 0$, then there exists a y such that $y^2 = 0$.

(d) If $2x + 1 = 5$, then either $x = 2$ or $x = 3$.

2.5.2 State the converse and contrapositive of each of the statements in Exercise 2.5.1.

2.5.3 Write the negation of each of the statements in Exercise 2.5.1.

2.5.4 Write each of the following statements in "if . . . , then . . . " form.

(a) It is necessary to score at least 90% on the test in order to receive an A.

(b) A sufficient condition for passing the test is a score of 50% or higher.

(c) You fail only if you score less than 50%.

(d) You succeed whenever you try hard.

2.5.5 State the converse and contrapositive of each of the statements in Exercise 2.5.4.

2.5.6 Write the negation of each of the statements in Exercise 2.5.4.

2.5.7 Restate each of the following assertions in "if . . . , then . . . " form.

(a) Perpendicular lines must intersect.

(b) Any two great circles on $\mathbb{S}^2$ intersect.

(c) Congruent triangles are similar.

(d) Every triangle has angle sum less than or equal to 180°.

2.5.8 Identify the hypothesis and conclusion of each of the following statements.

(a) I can take topology if I pass geometry.

(b) I get wet whenever it rains.

(c) A number is divisible by 4 only if it is even.

2.5.9 Restate using quantifiers.

(a) Every triangle has an angle sum of 180°.

(b) Some triangles have an angle sum of less than 180°.

(c) Not every triangle has angle sum 180°.

(d) Any two great circles on $\mathbb{S}^2$ intersect.

(e) If P is a point and ℓ is a line, then there is a line m such that P lies on m and m is perpendicular to ℓ.

2.5.10 Negate each of the following statements.

(a) There exists a model for incidence geometry in which the Euclidean Parallel Postulate holds.

(b) In every model for incidence geometry there are exactly seven points.

(c) Every triangle has an angle sum of 180°.

(d) Every triangle has an angle sum of less than 180°.

(e) It is hot and humid outside.

(f) My favorite color is red or green.
(g) If the sun shines, then we go hiking.
(h) All geometry students know how to write proofs.

2.5.11 Restate the Elliptic Parallel Postulate as a conditional statement.

2.5.12 Restate the Hyperbolic Parallel Postulate as a conditional statement.

2.5.13 Write the negation of the Elliptic Parallel Postulate.

2.5.14 Write the negation of the Hyperbolic Parallel Postulate.

2.5.15 Construct truth tables that illustrate De Morgan's Laws (page 31).

2.5.16 Construct a truth table which shows that the conditional statement $H \Rightarrow C$ is logically equivalent to $(\textit{not } H)$ *or* C. Then use one of De Morgan's Laws to conclude that *not* $(H \Rightarrow C)$ is logically equivalent to H *and* $(\textit{not } C)$.

2.5.17 Construct a truth table which shows directly that the negation of $H \Rightarrow C$ is logically equivalent to H *and* $(\textit{not } C)$.

2.5.18 State the Pythagorean Theorem in "if . . . , then . . . " form.

2.6 SOME THEOREMS FROM INCIDENCE GEOMETRY

We illustrate the lessons of the last section with several theorems and a proof from incidence geometry. The theorems in this section are theorems in incidence geometry, so their proofs are to be based on the three incidence axioms that were stated in Section 2.2. One of the hardest lessons to be learned in writing the proofs is that we may use only what is actually stated in the axioms, nothing more. Here, again, are the three axioms.

Incidence Axiom 1. For every pair of distinct points P and Q there exists exactly one line ℓ such that both P and Q lie on ℓ.

Incidence Axiom 2. For every line ℓ there exist at least two distinct points P and Q such that both P and Q lie on ℓ.

Incidence Axiom 3. There exist three points that do not all lie on any one line.

The first theorem was already used as an example earlier in the chapter. As explained in the last section, this theorem must be restated before it is ready for a proof. We will adopt the custom of formally restating theorems in if . . . then . . . form when necessary.

Definition 2.6.1. Two lines are said to *intersect* if there exists a point that lies on both lines.

Theorem 2.6.2. Lines that are not parallel intersect in one point.

Restatement. If ℓ and m are distinct nonparallel lines, then there exists a unique point P such that P lies on both ℓ and m.

Proof. Let ℓ and m be two lines such that $\ell \neq m$ and $\ell \nparallel m$ (hypothesis). We must prove two things: first, that there is a point that lies on both ℓ and m and, second, that there is only one such point.

There is a point P such that P lies on both ℓ and m (negation of the definition of parallel). Suppose there exists a second point Q, different from P, such that Q also lies